



**Contribution de l'épigénétique dans les
Dauermofifikationen et l'évolution adaptative chez le
parasite humain *Schistosoma mansoni* et le corail
tropical *Pocillopora damicornis***

David Roquis

► **To cite this version:**

David Roquis. Contribution de l'épigénétique dans les Dauermofifikationen et l'évolution adaptative chez le parasite humain *Schistosoma mansoni* et le corail tropical *Pocillopora damicornis*. Génétique. Université de Perpignan, 2015. Français. NNT : 2015PERP0037 . tel-01265805

HAL Id: tel-01265805

<https://theses.hal.science/tel-01265805>

Submitted on 1 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de
Docteur

Délivré par
UNIVERSITE DE PERPIGNAN VIA DOMITIA

Préparée au sein de l'école doctorale ED305

Et de l'unité de recherche UMR5244
Interactions Hôtes Pathogènes Environnements (IHPE)

Discipline : **Biologie**
Spécialité : **Épigénétique**

Présentée par **David ROQUIS**

**Contribution de l'épigénétique dans les
Dauermodifikationen et l'évolution adaptative chez le
parasite humain *Schistosoma mansoni* et le corail
tropical *Pocillopora damicornis***

Soutenue le 8 décembre 2015 devant le jury composé de

M. Étienne DANCHIN, DRCE, Université de Toulouse 3
M. Raymond PIERCE, DR2, Université de Lille 1
Mme. Marie MIROUZE, CR, Université de Montpellier
M. Olivier REY, CR, Université de Swansea
M. Guillaume MITTA, Professeur, Université de Perpignan
Mme Céline COSSEAU, MCF, Université de Perpignan
M. Christoph GRUNAU, Professeur, Université de Perpignan

Rapporteur
Rapporteur
Examineur
Examineur
Examineur
Directeur de thèse
Co-directeur de thèse

À Nané et Bon'pa

"Biology is the study of complicated things that have the appearance of having been designed with a purpose."

Richard Dawkins

Remerciements

Beaucoup m'ont permis de faire de ces dernières années une belle et enrichissante expérience de vie, autant sur le plan humain que scientifique. Je pense qu'il est logique de commencer par remercier mon laboratoire d'accueil, l'UMR 5244 IHPE. Ma thèse n'a jamais été un fardeau ou un calvaire, et je pense que c'est vraiment grâce à l'environnement de travail et l'ouverture d'esprit de mes collègues du laboratoire. Merci à vous tous pour cette ambiance, votre soutien quand j'en ai eu besoin, votre compagnie et nos nombreuses discussions. J'ai quelques mentions spéciales, et les premières vont à Nathalie Arancibia, Jérôme Boissier et Anne Rognon qui m'ont beaucoup aidé pour tout ce qui a trait à l'expérimentation animale. Je n'ai jamais été très à l'aise avec la manipulation des animaux vertébrés (c'était d'ailleurs ma plus grande réticence à faire une thèse dans ce laboratoire), et je n'aurais pas accompli grand chose sans votre aide patiente et compréhensive. Nathalie, avec qui j'ai fait les infestations et découvert la subtile différence entre anesthésie et euthanasie (les produits se ressemblent, mais n'ont pas tout à fait les mêmes effets...). Jérôme, pour toutes les perfusions et pour m'avoir parlé des greffes de sporocystes, et Anne pour avoir ressorti ton matériel artisanal et ton art de microchirurgie molluscale. Merci à Jean-François Allienne, la fée des paillasses, pour ton rôle incontournable dans la logistique expérimentale. J'ai également une pensée pour Benjamin Gourbal, acteur central de mon « dépuçage ». Merci pour cette action antiparasitaire, mais aussi pour nos échanges sur les interactions hôtes/pathogènes et la notion de compatibilité, qui m'ont beaucoup fait réfléchir au design expérimental de ma thèse. Cristian Chaparro, merci de m'avoir soulagé dans la gestion de la bioinformatique. Merci également à Guillaume Mitta, pour avoir cru en moi, m'avoir aidé dans mes démarches et préparé au concours de l'école doctorale. Petit clin d'œil pour Cécile St-Béat; pour ton petit côté pétillant qui met toujours de bonne humeur. Je finis avec une pensée pour feu Bernard Dejean. Personne n'a remplacé ta verve et ta poésie.

En dehors du laboratoire, mais toujours dans le contexte de travail, je remercie les membres de mes comités de suivi de thèse, pour leur regard critique et conseils : Marie Mirouze, Guilherme Oliveira (grâce à qui je me suis intéressé à la bioinformatique) et Bernt Schuttengrubber. Merci également à Aaron Taudt, Maria Colomé-Tatché et Frank Johannes pour m'avoir accueilli à Gronningen et fait progresser mes analyses bioinformatiques.

Évidemment, mon allégresse professionnelle aurait été tout autre sans mes compagnons du bureau, indomptables doctorants, bravant promiscuité, intenses conditions climatiques, cacophonie, et insectes hématophages pour servir la Science avec dévouement. Marion Picard, Sara Fneich, c'est pour votre accueil, à cette lointaine époque où ce bureau nous paraissait si vaste. Je garde un bon souvenir de nos discussions tardives. Rodolfo Rondon, tu parles peu, mais tu parles juste. Je suis content que tu nous aies rejoint en cours de route, et merci d'avoir supporté mes bavardages. Silvain Pinaud et Anaïs Portet, un peu de sang frais qui a apporté beaucoup de dynamisme à notre groupe. Nele Boon, pour ton ouverture d'esprit et ta vision éclairée sur la schistosomiase. Kelly Brener.... Tu mériterais une page de remerciement à toi toute seule (et tu le sais)... Malheureusement, j'avais déjà fait cette promesse à quelqu'un d'autre il y a fort longtemps. Je pense de toute manière que

notre amitié sincère et symbiotique se passe de mots. Et non, je ne dis pas ça par manque d'inspiration. Finalement, Grosse Pétouille, je ne t'oublie pas, fière et gracieuse Aiptasia pallida trônant dans son verre à cognac sur mon bureau. Tu ne liras jamais ces mots (puisque tu es une anémone de mer), mais j'espère que tu traîneras à mes côtés suffisamment longtemps pour que je puisse te léguer à mes étudiants à ma retraite.

Dans le cadre universitaire, certains de mes enseignants ou de mes anciens étudiants m'ont aussi offert la chance de m'enrichir à leur contact. Nathalie Picault, bien sûr, pour tes nombreux apports dans le cheminement de mes études, notre covoiturage quotidien, ton dynamisme et toutes nos conversations scientifiques et moins scientifiques. Constance Tisserant, pour ton aide précieuse sur l'épigénétique des Daphnies, et Ariadna Picart Picolo pour ton travail efficace sur le méthylome de Puccinellia (et pour ta contribution pâtissière à la soirée doctorante la plus distrayante à laquelle j'ai participé). Ninon de Mecquenem, pour m'avoir offert le plus beau compliment que j'ai eu sur ma manière d'enseigner. Ce jour là, tu m'as vraiment convaincu que j'avais trouvé ma vocation. Je suis aussi reconnaissant envers mes anciens enseignants de Chimie-Biologie du CEGEP Ahuntsic. L'excellente formation technique et théorique que j'ai reçue de vous a été un atout indiscutable de ma vie professionnelle et de ma reprise d'études. Je pense particulièrement à Lucie Brouillette et Sylvain Beausoleil, qui m'ont offert plus que de la simple connaissance scientifique. Mon dernier mot va à Laurent Derepentigny, professeur de biologie à la polyvalente Louis-Philippe Paré, qui a allumé la flamme de la biologie en moi, il y a déjà 20 ans. Nous nous sommes perdus de vue, mais j'espère que tu liras ces lignes un jour. Tu y es pour beaucoup.

Un grand remerciement à tous mes amis qui m'ont épaulé pendant ces trois ans. Les littéraires, qui m'ont permis de garder l'esprit ouvert à autre chose que la biologie : Marie David, à qui je dois une fière chandelle pour la relecture de ce manuscrit, Julien Ortega, Marie Cremades (oui je sais, tu es archéologue, mais bientôt tu seras littéraire), Julien Bougrelle et Angélique Bertot. Merci à ma compagne culturelle et gastronome, Marianne Miguët, qui m'a permis de me changer les idées à de nombreuses reprises. Marie Suez, j'espérais soutenir en même temps que toi, mais tu m'as battue de deux mois et demi. Un mot aussi pour les amis lointains, restés à Montréal, en particulier Daniel Boileau, Corine Zotti Olivier Bernèche et Jocelyne Ayotte. J'ai beaucoup grandi grâce à vous quatre.

Pierre Lepage, je t'avais promis une page complète de remerciement si un jour j'écrivais une thèse. Je vais être parjure, mais je crois que tu n'auras qu'un paragraphe. C'est quand même pas mal, tu es la seule personne à avoir eu son paragraphe dévoué. Je voudrais te remercier pour tout ce que tu as fait pour moi durant ces quatorze dernières années. C'est toi qui m'as offert ma chance en science en m'offrant un poste à Génome Québec, et c'est toi qui m'as poussé vers la sortie quelques années plus tard pour que je retourne aux études. Je pense que je n'aurais jamais mis les pieds à l'université si je n'avais pas croisé ton chemin. Tes critiques justes m'ont toujours permis de me remettre en question au bon moment. Et merci aussi d'avoir pris le temps de me trouver toute la littérature scientifique à laquelle je n'avais pas accès.

J'embrasse aussi ma famille, en particulier mes parents, Marie-Christine Delaferme et Jean-Michel Bloch, mon parrain Jean-Claude Delaferme, et mes grands-parents, Bon'pa et Nané, pour tous vos encouragements et votre soutien depuis mon arrivée en France. Je t'embrasse aussi Alexandra Barrat,, mais amoureusement. Merci de m'avoir supporté pendant ces trois années, surtout pendant ma période de rédaction, et pour tout ton amour (et non, ce n'est pas parce que tu as moins de lignes de remerciements que Grosse Pétouille que je t'aime moins). Cette thèse vous est dédiée.

Je termine avec les plus importants, Céline Cosseau et Christoph Grunau. Vraiment, je n'aurais pu souhaiter un meilleur encadrement. Merci de vous être tant investi pour moi, pour m'avoir guidé sans jamais me diriger, pour m'avoir soutenu, fait confiance et inspiré. Voilà déjà cinq ans que nous travaillons ensemble, depuis le petit Axe 3 jusqu'à la moins petite équipe BMC, et c'est un peu perturbant de réaliser que c'est une page qui se tourne. Céline, merci pour le temps que tu as su me consacrer malgré ton emploi du temps compliqué. Christoph, tu as toujours su trouver les mots justes, au moment où j'avais besoin de les entendre. Merci de m'avoir fait découvrir une autre vision de la science et de la pédagogie. Tu as été une figure paternelle pour moi. J'espère un jour être en mesure de retransmettre tout ce que vous m'avez apporté à d'autres étudiants.

La rédaction de ce manuscrit a bénéficié de l'inspiration musicale de Fela Anikulapo Kuti et de l'inspiration maltée de Stéphanie Altermatt

Cette version corrigée du manuscrit ayant été produite post-soutenance, je tiens à sincèrement remercier les membres du jury (Marie Mirouze, Étienne Danchin, Raymond Pierce, Olivier Ray et Guillaume Mitta) pour leurs commentaires et appréciations qui m'ont permis d'améliorer mon travail et ma réflexion scientifique, mais également pour leurs contributions à cette merveilleuse expérience de vie qu'aura été ma soutenance du 8 décembre 2015.

J'en profite également pour exprimer ma gratitude envers Julie et Ludovic Bouaziz, du restaurant Sabaïdee (Saint Laurent de la Salanque), pour l'organisation sans faille du pot suivant la soutenance (et ce, malgré la blessure de Ludovic) ainsi que Laurent Drajkowski, du restaurant le Saint-Jean (Perpignan), pour nous avoir exceptionnellement bien reçu dans la soirée, et avoir été particulièrement accommodant malgré l'organisation tardive des festivités.

Et finalement, je consacre le tout dernier merci de ce manuscrit à Manon Fallet, pour son œil d'aigle dans la correction des dernières fautes et coquilles de cette version définitive du manuscrit.

TABLE DES MATIÈRES

PRÉAMBULE	1
CHAPITRE I: INTRODUCTION GÉNÉRALE	5
I.1. ÉVOLUTION, ADAPTATION ET HÉRITABILITÉ	7
I.1.1. L'ADAPTATION	7
I.1.2. PLASTICITÉ PHÉNOTYPIQUE ET NORME DE RÉACTION	7
I.1.3. L'ÉVOLUTION ADAPTATIVE, UN CONCEPT CHANGEANT DANS L'HISTOIRE	9
I.1.3.1. Antiquité et théories philosophiques précoces	9
I.1.3.2. Moyen-âge, une vision théologique	12
I.1.3.3. XVII et XVIII ^{ème} siècles, essor de la taxonomie	12
I.1.3.4. Lamarck, première théorie scientifique	13
I.1.3.5. Darwin et la sélection naturelle	14
I.1.3.6. Les expériences de Weismann et la chute du transformisme	15
I.1.3.7. Émergence de la génétique	16
I.1.3.8. Théorie synthétique de l'évolution	17
I.1.3.9. Progrès moléculaires et génocentrisme	18
I.1.3.10. Importance des interactions	18
I.1.4. HÉRÉDITÉ DES CARACTÈRES ACQUIS	20
I.1.4.1. Victor Jollos et les Dauermodifikationen	20
I.1.4.2. Exemples d'hérédité non-mendélienne	21
I.1.4.3. Origines possibles des Dauermodifikationen	24
I.2. L'ÉPIGÉNÉTIQUE	26
I.2.1. BREF HISTORIQUE DE LA DISCIPLINE	26
I.2.1.1. Embryologie, épigénèse et préformationnisme	26
I.2.1.2. Le paysage épigénétique de Waddington	27
I.2.1.2. L'épigénétique moderne	29
I.2.2. PORTEURS D'INFORMATION ÉPIGÉNÉTIQUE	31
I.2.2.1. Méthylation des cytosines	32
I.2.2.2. Autres modifications chimiques des nucléotides	36
I.2.2.3. Modification des histones et position des nucléosomes	38
I.2.2.4. ARNs non-codants	42
I.2.2.5. Localisation nucléaire des chromosomes	43
I.2.3. LIENS ENTRE DAUERMODIFICATIONS ET ÉPIGÉNÉTIQUES	45
I.2.3.1. Caractéristiques d'un modèle d'étude Dauermodifikationen/épigénétique	46
I.3. SCHISTOSOMA MANSONI, UN PARASITE HUMAIN	47
I.3.1. DESCRIPTION DU MODÈLE	47
I.3.1.1. Cycle de vie	47
I.3.1.2. La bilharziose, maladie tropicale négligée aux lourdes conséquences	49
I.3.1.3. Plasticité environnementale, de possibles Dauermodifikationen?	51
I.3.4. ÉTAT DE L'ART DES CONNAISSANCES ÉPIGÉNÉTIQUES CHEZ S. MANSONI	52
I.3.5. PRÉCÉDENTES EXPÉRIENCES ET RÉSULTATS	54
I.3.5.1. Impact de la souche d'hôte intermédiaire sur le développement parasitaire	54
I.3.5.2. Résistance induite à un schistosomicide	57
I.4. LE CORAIL TROPICAL POCILLOPORA DAMICORNIS	58
I.4.1 DESCRIPTION DU MODÈLE	58
I.4.1.1. Caractéristiques biologiques et cycle de vie	58
I.4.1.2. Distribution géographique	60
I.4.1.3. Importance écologique et économique	61
I.4.1.4 Plasticité environnementale : thermotolérance	61
I.4.2. ÉTAT DE L'ART DES CONNAISSANCES ÉPIGÉNÉTIQUES CHEZ P. DAMICORNIS	63
I.5. OBJECTIFS DE LA THÈSE	64
I.5.1. HYPOTHÈSES ET OBJECTIFS GÉNÉRAUX	64
I.5.2. OBJECTIFS SPÉCIFIQUES À SCHISTOSOMA MANSONI	67
I.6.3. OBJECTIFS SPÉCIFIQUES À POCILLOPORA DAMICORNIS	68

CHAPITRE II: CARACTÉRISATION DE LA STRUCTURE CHROMATINIENNE DE <i>S. MANSONI</i>/.....	69
II.1. MISE EN CONTEXTE.....	71
II.2. PUBLICATION ASSOCIÉE	72
CHAPITRE III: RÉSISTANCE À UN SCHISTOSOMICIDE ET IMPACT SUR L'ÉPIGÉNOME DE <i>S. MANSONI</i>. 99	
III.1. MISE EN CONTEXTE.....	101
III.2. PUBLICATION ASSOCIÉE	103
CHAPITRE IV: EFFET DE LA SOUCHE D'HÔTE INTERMÉDIAIRE SUR L'ÉPIGÉNOME DE <i>S. MANSONI</i>/....	121
IV.1. MISE EN CONTEXTE	123
IV.2. PUBLICATION ASSOCIÉE.....	125
IV.3. RÉSULTATS ET DISCUSSION SUPPLÉMENTAIRES NON PUBLIÉS	171
IV.3.1. EFFET DE LA SOUCHE DE <i>B. GLABRATA</i> SUR LE SUCCÈS DE GREFFE	171
IV.3.2. EFFET DU GÉNOTYPE DU PARASITE SUR L'INTENSITÉ DE L'INFECTION CHEZ LA SOURIS	173
IV.4. PERSPECTIVES	173
CHAPITRE V: TECHNIQUES D'ÉTUDE DE L'ÉPIGÉNOME DE <i>POCILLOPORA DAMICORNIS</i>	175
V.1. MISE EN CONTEXTE	177
V.2. PUBLICATION ASSOCIÉE.....	178
V.3. RÉSULTATS ET DISCUSSION SUPPLÉMENTAIRES NON PUBLIÉS.....	193
V.3.1. OPTIMISATION DE LA TECHNIQUE D'IMMUNOPRÉCIPITATION DE CHROMATINE	193
V.3.2. COMPARAISON DES MÉTHYLOMES ENTRE CORAUX THERMOTOLÉRANTS ET SENSIBLES	195
CHAPITRE VI: DISCUSSION	197
VI.1. RAPPEL DES PRINCIPAUX RÉSULTATS	199
VI.1.1. DESCRIPTION DES ÉPIGÉNOMES DE <i>S. MANSONI</i> ET <i>P. DAMICORNIS</i>	199
VI.1.1.1. Structure chromatinienne différente entre <i>S. mansoni</i> cercaires et adultes	199
VI.1.1.2. Premier méthylome d'un corail scléractiniaire.....	200
VI.1.2. EFFET DES CHANGEMENTS ENVIRONNEMENTAUX SUR L'ÉPIGÉNOME.....	201
VI.1.2.1. Résistance à l'hycanphone associée à des épimutations.....	201
VI.1.2.2. Trois types d'épimutations lors des greffes de sporocystes.....	203
VI.1.2.3. Altération du méthylome chez les coraux ayant subi un blanchissement.....	204
VI.1.3. HÉRITABILITÉ DES ÉPIMUTATIONS	204
VI.2. CONSIDÉRATIONS EXPÉRIMENTALES	205
VI.2.1. IMPORTANCE GÉNÉTIQUE ET ÉPIGÉNÉTIQUE DE L'ÉTUDE DES DAUERMODIFIKATIONS	205
VI.2.2. CONTRIBUTIONS RESPECTIVES DE L'ÉPIGÉNÉTIQUE ET LA GÉNÉTIQUE AU PHÉNOTYPE	207
VI.2.3. IMPORTANCE DE L'ANALYSE INDIVIDUELLE DES RÉPLICATS BIOLOGIQUE.....	208
VI.2.4. MESURE ET QUANTIFICATION DE LA VARIABILITÉ OBSERVABLE	209
VI.3. ORIGINE DE LA VARIABILITÉ ÉPIGÉNÉTIQUE	210
VI.4. RÔLE ÉVOLUTIF DE LA VARIABILITÉ ÉPIGÉNÉTIQUE.....	213
CHAPITRE VII: CONCLUSION ET PERSPECTIVES.....	219
CHAPITRE VIII: RÉFÉRENCES BIBLIOGRAPHIQUES.....	223

CHAPITRE IX: ANNEXES	245
IX.1. ANNEXE 1 : PUBLICATION SUPPLÉMENTAIRE 1.....	247
IX.1.1. MISE EN CONTEXTE	247
IX.1.2. PUBLICATION ASSOCIÉE	247
IX.2. ANNEXE 2 : PUBLICATION SUPPLÉMENTAIRE 2.....	261
IX.2.1. MISE EN CONTEXTE	261
IX.2.2. PUBLICATION ASSOCIÉE	261
IX.3. ANNEXE 3 : PROTOCOLES DE CHIP TESTÉS SUR <i>P. DAMICORNIS</i>.....	273
IX.4. ANNEXE 4 : DIAPOSITIVES DE SOUTENANCE	275
IX.5. ANNEXE 5 : DOCUMENTS ADMINISTRATIFS RELATIFS À LA THÈSE.....	289
IX.6. ANNEXE 6 : LISTE DES CORRECTIONS DU MANUSCRIT DE THÈSE	305

PRÉAMBULE

J'ai commencé mes études en biologie, au CEGEP Ahuntsic, à Montréal, à la fin du XX^{ème} siècle (àïe, les années tournent...), au sein d'une formation technique en biotechnologie. À cette époque, les choses étaient simples. Un gène, une protéine, une fonction. Tout ce qui n'était pas apparenté de près ou de loin à un gène était de l'ADN poubelle, sans véritable fonction. Quelques études avaient attribué par-ci par-là certains rôles à des séquences non géniques, mais on m'enseigna avec conviction qu'il s'agissait des rares exceptions confirmant la règle. L'évolution? Une histoire de mutations aléatoires suivie par la sélection des individus auxquels les mutations offraient de meilleurs atouts. J'étais bien conscient qu'il existait quelques subtilités additionnelles, mais dans l'ensemble, ça se résumait à ça. Il restait encore quelques mystères à éclaircir, mais le gros du terrain était débroussaillé, et la vue d'ensemble assez claire.

J'ai par la suite travaillé comme technicien dans un centre de génétique humaine, de 2002 à 2009 (avec une parenthèse d'un an en Thaïlande sur la génétique du riz). La publication du génome humain (avec beaucoup moins de gènes qu'initialement pensé) et les premières importances fonctionnelles que prenaient certains éléments de « *junk DNA* » ont complexifié quelque peu ma perception de la génétique, mais jamais sans véritablement remettre en cause ses fondements et son rôle central dans l'évolution.

En 2009, alors que j'avais une vie tout à fait agréable, j'ai eu l'idée un peu étrange de retourner à l'université à temps plein. Cela faisait quelques temps que je faisais des ateliers scientifiques avec des lycées, et l'idée de me reconverter en enseignant de biologie me tentait bien. J'avais commencé une licence en biochimie à temps partiel en 2004, et j'ai eu l'idée de la terminer à l'Université de Perpignan (pour des considérations à la base purement financières). Finalement, j'ai aimé ça, je me suis lancé dans un Master et j'ai laissé tomber l'idée d'être professeur au lycée (maître des conférences à l'université offrant la possibilité de partager une bière avec mes futurs étudiants). C'est à ce moment-là que j'ai eu mon premier vrai contact avec l'épigénétique, une discipline dont l'enseignement récent paraissait en vogue. S'il me semblait clair qu'elle jouait un rôle dans l'acclimatation d'un individu à son environnement, je restais sceptique quant à son implication dans l'hérédité et l'évolution. Il me semblait un peu inconcevable que deux cents ans de travaux scientifiques aient pu rater quelque chose qui paraissait avoir des répercussions importantes en terme d'évolution adaptative.

À la fin de mon second stage au laboratoire IHPE, mes encadrants ont mentionné la possibilité de faire un doctorat. Le sujet proposé se basait sur des expériences précédentes mettant en avant une héritabilité de certains caractères acquis pour essayer d'en déterminer les causes moléculaires (présumées d'origine épigénétique). Je n'ai rien dit à mes encadrants à ce moment-là pour ne pas leur faire de peine (ils avaient l'air de tellement croire en cette épigénétique!), mais cette pseudo-héritabilité douteuse de certains traits était forcément dans l'ombre de variations génétiques qui n'avaient pas été identifiées. Peut-être même simplement des artéfacts expérimentaux.

Essayant tout de même de ne pas être trop critique, j'ai à l'époque épluché la littérature scientifique avec pour objectif de recenser (non exhaustivement, mais de manière la plus complète possible) des études mettant en évidence une héritabilité de caractères adaptatifs chez des organismes pour lesquels génomes et épigénomes auraient été analysés en parallèle. Il y a 6 ans, je ne me rappelle pas avoir trouvé un article de la sorte (à part peut-être chez les végétaux), mais j'ai quand même lu quelques papiers dont les résultats et conclusions m'ont intrigués, et donnés la curiosité de continuer à chercher un peu plus loin. Dans cette optique, j'ai essayé de reconstruire le cheminement du raisonnement scientifique qui avait mené à toutes les certitudes que l'on m'avait inculqué par rapport à l'évolution et l'hérédité, et qui me semblaient difficile à mettre en cause. Mes études auront été marquées par un manque critique de cours d'épistémologie.

Je n'ai pas retourné ma veste du jour au lendemain, et je n'ai pas non plus rejeté les fondements que j'avais de la pensée évolutionniste, mais il m'a semblé que des aspects de ces dogmes avaient été construits sur des raisons plus historiques et contextuelles que purement scientifiques. Finalement, se pourrait-il qu'il existe des subtilités ignorées ou rejetées parce qu'elles ne rentraient pas dans le moule établi? C'est cet aspect historique (en particulier August Weismann et ses souris, dont je parlerai plus tard) qui m'a convaincu de remettre quelque peu mes connaissances en causes et de me lancer dans une thèse. C'est pour cela que j'ai essayé lui faire une place de choix à l'histoire des sciences dans ce manuscrit. Ma thèse s'achève, et j'ignore si je suis devenu un bon scientifique, mais je pense avoir appris à douter et à garder l'esprit ouvert, et juste pour cela, le jeu en valait la chandelle. Je comprends mieux ce que voulait dire Theodosius Dobzhansky par « *Rien en biologie n'a de sens, si ce n'est à la lumière de l'évolution* ». Bonne lecture!

CHAPITRE I

INTRODUCTION GÉNÉRALE

I.1. ÉVOLUTION, ADAPTATION ET HÉRITABILITÉ

I.1.1. L'adaptation

L'environnement joue un rôle central dans les dynamiques évolutives. Ce dernier exerce des pressions, abiotiques (température, conditions physico-chimiques) et biotiques (influence d'autres organismes vivants partageant le même environnement) sur les individus et espèces qui le peuplent. Pour être en mesure de s'installer et proliférer dans un environnement, un organisme donné doit y être adapté. On peut définir une adaptation, comme un caractère ayant un rôle fonctionnel au sein d'un trait d'histoire de vie et qui accroît le succès reproducteur (*i.e.* sa fitness) des individus qui le possèdent. L'adaptation peut aussi faire référence aux processus par lesquels les individus d'une population tirent un meilleur profit d'une composante de leur environnement via la modification d'un de leur caractère (ayant un impact sur la survie ou la reproduction) (Futuyma 2013). Theodosius Dobzhansky définit trois niveaux: **(I)** l'adaptation, un processus évolutif par lequel un organisme devient plus performant à vivre dans son ou ses habitats; **(II)** l'adaptabilité, le degré selon lequel un organisme est capable de vivre et se reproduire dans un habitat donné; **(III)** un trait adaptatif, un aspect du profil développemental d'organisme qui permet ou améliore la probabilité de survie et de reproduction de cet organisme (Dobzhansky & Dobzhansky 1970). L'adaptation est un des deux processus, avec la spéciation causée par un isolement géographique, qui permet d'expliquer la diversité biologique des organismes (Mayr 1982).

I.1.2. Plasticité phénotypique et norme de réaction

Une adaptation n'est pas forcément reliée à un unique phénotype. L'environnement pouvant être fluctuant durant la durée de vie d'une espèce, un individu en mesure de répondre efficacement à plusieurs conditions sera favorisé et aura une fitness accrue. **(Figure 1)**. Whitman & Agrawal (2009) la définisse comme la capacité d'un organisme à exhiber différents variants phénotypiques à partir d'un seul génotype, en réponse aux fluctuations environnementales. Elle est considérée comme une adaptation évolutive si elle permet de répondre à des fluctuations plus ou moins prévisibles de l'environnement (ex : changement de saisons) et peut être soumise à la sélection naturelle (DeJong *et al.* 2001).

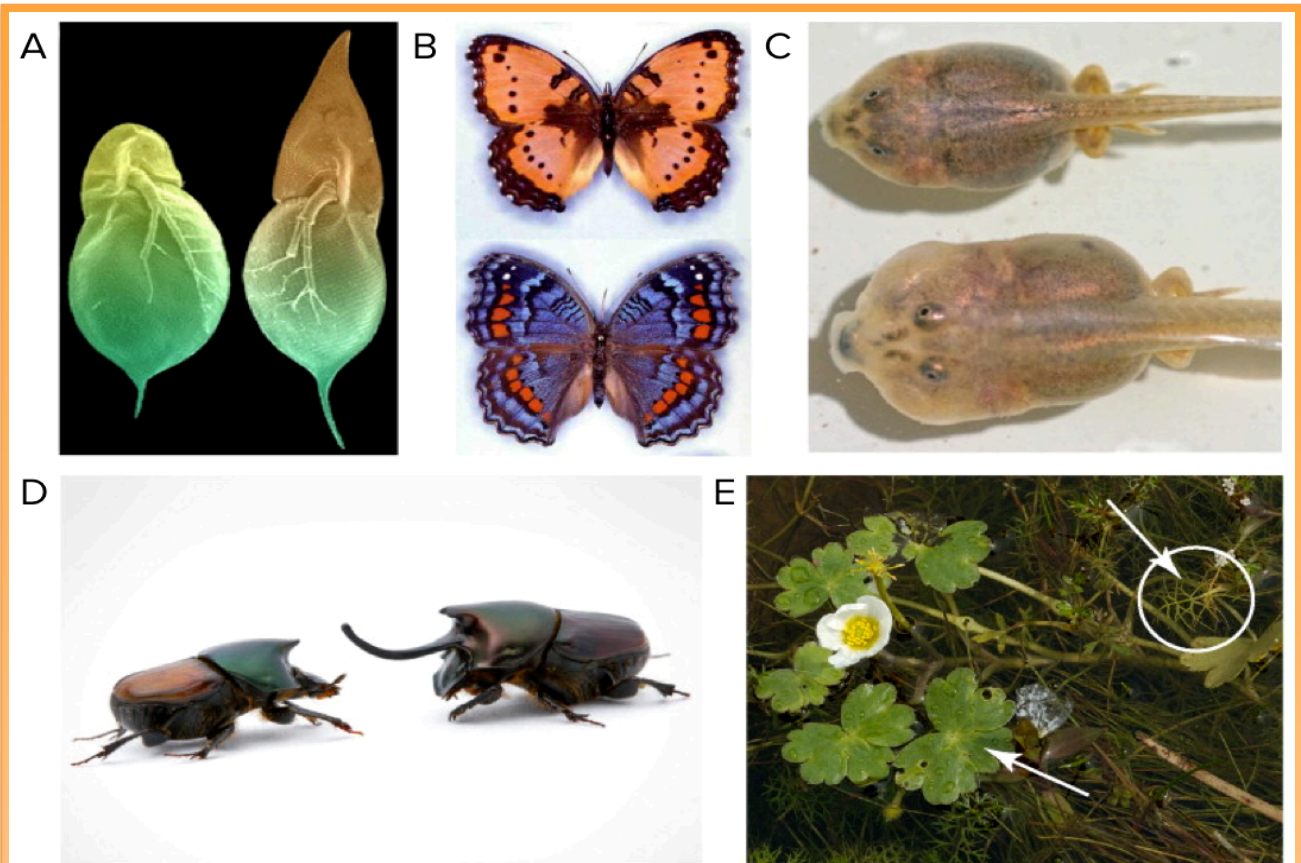
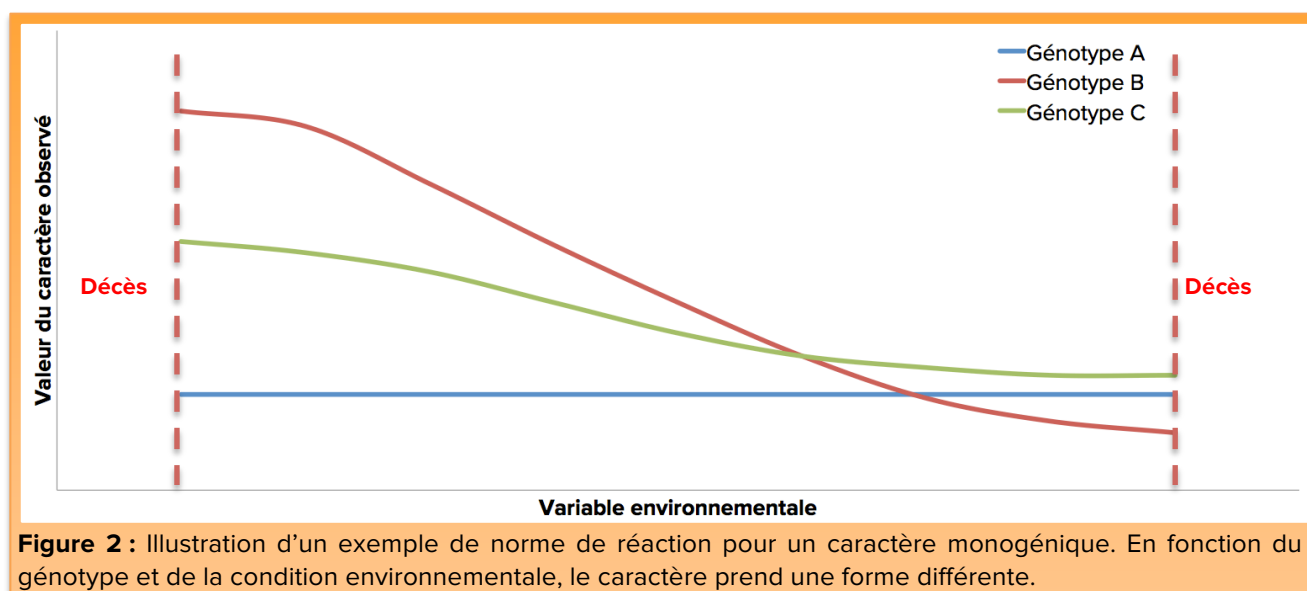


Figure 1: Exemples de plasticité phénotypiques chez différents organismes. **(A)** *Daphnia cucullata* en condition normale (gauche) et en présence d'un prédateur (droite). **(B)** Formes saisonnières du papillon *Precis octavia* (saison humide en haut, saison sèche en bas). **(C)** Formes omnivore et carnivore du têtard de *Spea multiplicata*. **(D)** Formes à courte et longue corne du coléoptère *Onthophagus nigriventris*. **(E)** Feuilles aériennes et aquatiques de *Anunculus aquatilis*. Figure tirée de Pfennig et al. (2010) .

La plasticité phénotypique est au centre du phénomène d'acclimatation (ajustement d'un individu à des variations environnementales graduelles). On considère qu'elle a en général une importance plus grande chez les organismes sessiles, n'ayant pas l'opportunité de se déplacer pour échapper à des conditions défavorables (Whitman & Agrawal 2009). Il existe différents types de plasticité phénotypique; certaines sont réversibles (ex : bronzage chez l'être humain) ou irréversible (effet de la diète sur le phénotype de certains insectes, ou déterminisme sexuel de certains reptiles en fonction de la température d'incubation des œufs) (Whitman & Agrawal 2009). On utilise généralement la norme de réaction qui décrit le profil d'apparition d'un phénotype en fonction du génotype de l'individu et des conditions environnementales (**Figure 2**). C'est la fonction qui relie les environnements auxquels un génotype particulier peut être confronté et les phénotypes qui peuvent être produits par ce génotype. Ainsi, chaque individu au sein d'une population possède sa propre norme de réaction pour un caractère donné (Sarkar 1999).



I.1.3. L'évolution adaptative, un concept changeant dans l'histoire

Les origines biologiques et évolutives de l'adaptation et la plasticité phénotypique ont été historiquement très débattues, et certaines théories restent encore controversées de nos jours. Dans le cadre de l'adaptation par exemple, deux grandes écoles de pensées se sont confrontées au cours des trois derniers siècles. La première, issue des théories de Jean-Baptiste de Lamarck, la décrit comme un processus continu par lequel un organisme modifie ses caractères en réponse à l'environnement, et transmet ses caractères modifiés à la descendance (de Lamarck 1815). Cette vision laisse entrevoir une sorte de plasticité phénotypique, puisque les changements sont induits par la pression environnementale, mais avec la différence qu'ils ne sont *a priori* pas réversibles et directement transmis à la progéniture. Cette idée sera considérée obsolète avec l'avènement de la théorie synthétique de l'évolution, pour laquelle les traits adaptatifs apparaissent par mutations aléatoires, sont maintenus et évoluent par la pression de la sélection naturelle (Huxley 1943). Un bref récapitulatif de l'histoire de la pensée évolutive permet d'en comprendre les principales notions ainsi que le rôle clef de certaines expériences dans la vision actuelle.

I.1.3.1. Antiquité et théories philosophiques précoces

Les premières traces d'un questionnement sur l'origine de la vie apparaissent chez les philosophes grecs de l'Antiquité. Un des plus précoce, Anaximandre de Millet (610 av. J.-C. – 546 av. J.-C.) propose une origine aquatique des animaux. Il suppose l'existence

d'une période durant laquelle la Terre fut recouverte par les océans, et que les premiers animaux terrestres seraient apparus lors de l'assèchement de grands territoires. Il suggère que l'être humain est le produit de la descendance d'une autre espèce, argumentant que l'espèce humaine nécessite une trop longue période de soin à l'enfance (comparativement aux autres animaux) (Kirk *et al.* 1983). Sa vision était particulièrement en avance sur son époque, mais ne sera que peu diffusée. Presque un siècle plus tard, Empédocle (490 av. J.-C. – 430 av. J.-C.) affirme que les organismes vivants actuels sont des combinaisons d'organismes ancestraux aux caractères plus simples (Kirk *et al.* 1983). Ces premières notions d'évolution ne seront pas reprises par des philosophes plus influents, tels que Platon (428 av. J.-C. – 348 av. J.-C.) et Aristote (384 av. J.-C. – 322 av. J.-C.), dont la doctrine essentialiste place l'origine des êtres vivants comme étant d'essence divine. Ils sont conçus avec une fonction et un rôle préétabli, et immuable (Mayr 1982). On doit à Aristote la *scala naturæ*, « l'échelle de la vie » (**Figure 3 A**), une première tentative de classification des êtres vivants en fonction de leur apparente complexité phénotypique (Singer 1931). Cette vision était très lourdement biaisée par l'idée qu'une plus grande complexité de caractères signifiait une plus grande proximité avec le divin. Cette représentation erronée est cependant restée d'actualité. Elle fait aujourd'hui figure d'une image d'Épinal de la culture populaire, illustrant (à tort) les théories d'évolutions darwinienne (**Figure 3 B et 3 C**). La « marche du progrès » illustrée par Francis Clark Howell (1925-2007), en est sa version la plus célèbre (**Figure 3 B et 3C**). Une perception différente émerge d'Épicure (341 av. J.-C. – 270 av. J.-C.), qui présuppose une origine spontanée à l'apparition des êtres vivants; la plupart de ces organismes arrivant ensuite à survivre et à se reproduire (Mayr 1982). Cette notion n'inclut pas d'évolution, mais anticipe une sorte de sélection naturelle. Quelques siècles plus tard, Augustin d'Hippone (354-430) est le dernier philosophe antique à renouer explicitement avec l'idée d'évolution. Il n'accepte que partiellement la vision théologique de création divine prévalant à l'époque, et émet l'hypothèse que les êtres vivants sont conçus avec un certain degré de potentialité, et qu'ils peuvent se transformer progressivement avec le temps (Taylor 1982).

Il est important de noter que toutes ces théories précoces sont cantonnées au domaine de la philosophie, car aucun raisonnement ou expérimentation scientifique n'a été la stricte base de leurs constructions.



Figure 3 : Représentation de la *scala naturæ* (A) selon la description d'Aristote. Figure tirée de Singer (1931) (B) Capture d'écran des résultats images de Google.fr avec le mot clef « Évolution », le 1^{er} octobre 2015. On remarque que les 20 premiers résultats sont des variations contemporaines de la *scala naturæ* (ici majoritairement sous la forme de la marche du progrès de Francis Clark Howell). (C) Timbres commémoratifs des travaux de Charles Darwin émis par plusieurs pays (Micronésie, Bosnie-Herzégovine, Italie, Grèce, États-Unis et Royaume-Uni). À deux exceptions près (Bosnie-Herzégovine et États-Unis), ils ont été émis à l'occasion du bicentenaire de la naissance de Darwin (2009). Ils sont tous associés, à tort, à la *scala naturæ* (une fois encore sous la forme de la marche du progrès de Francis Clark Howell). Ces exemples montrent que cette conception est toujours présente dans la culture populaire et injustement associée au concept d'évolution, en particulier au raisonnement darwinien.

I.1.3.2. Moyen-âge, une vision théologique

La période médiévale correspond à un retour en occident de l'œuvre des philosophes antiques, dont une partie des théories et conceptions sera incorporée dans le corpus religieux. La *scala naturæ* d'Aristote y connaît un retour en force et se retrouve consolidée par une version christianisée dans laquelle toutes les êtres vivants ont des positions uniques se succédant (il ne manque aucun maillon) dans l'échelle de la vie. L'être humain se trouve bien entendu à l'extrémité, juste avant le divin (Lovejoy & Stanlis 2011). À l'image de la perception antique, l'organisation des maillons de la chaîne se fait en fonction de la complexité apparente des organismes (qui n'est nullement un reflet de la véritable complexité biologique). Les espèces ont été créées telles quelles et ne changent pas. Thomas d'Aquin (1225-1274) fait preuve d'une plus grande ouverture théologique et poursuit des réflexions similaires à Augustin d'Hippone, à savoir qu'il est envisageable que la vie puisse évoluer au travers de mécanismes naturels (Carroll 2000).

Si cette époque n'est pas propice aux développements de la pensée évolutionniste en Europe, au Moyen-Orient, quelques penseurs travailleront sur cette thématique. Al-Jāhiz (776-868) décrit des mécanismes d'interactions entre espèces similaires aux bases de la sélection naturelle (Zirkle 1941). Nasīr al-Dīn Tūsī (1201-1274) évoque une origine simiesque à l'espèce humaine, qui serait apparue par des processus de perfectionnement de caractères innés sous l'influence de l'environnement. Il propose que les organismes pouvant acquérir plus facilement de nouveaux caractères soient favorisés (Zirkle 1941).

I.2.3.3. XVII^{ème} et XVIII^{ème} siècles, essor de la taxonomie

La Renaissance se caractérise par l'apparition formelle du terme « évolution » dans la langue anglaise, qui dérive du latin *evolutio* et signifiant « action de dérouler ». Le mot fait alors initialement référence aux processus de développement embryonnaire et sera utilisé en relation à la progression biologique des espèces par Charles Bonnet (1720-1793) en 1762 (Goodrum 2002). La plupart des théories sur les origines de la vie au XVII^{ème} siècle restent imprégnées du rôle de l'intervention divine et l'évolution y est définie comme un processus non pas biologique, mais spirituel (Schelling 1993). John Ray (1627-1705) est le premier à donner une définition biologique d'espèce, ancrée dans une conception fixiste de la vie, selon laquelle il n'y a ni dérive, ni changement des espèces avec le temps (Wilkins 2006).

Cette vision immuable des organismes vivants sera partagée par de nombreux naturalistes de la période des Lumières. Carl von Linné (1707-1778) instaure le premier système moderne de taxonomie, dans l'objectif de mieux comprendre la hiérarchie divine des espèces (toujours en référence à la *scala naturæ*). Il n'établira cependant pas de liens phylogéniques entre les espèces, bien qu'il les regroupera en catégories par similarité de caractères (Reece 2011). Allant légèrement à l'encontre des tendances de son époque, Georges-Louis Leclerc, comte de Buffon (1707-1788) suggère que les espèces d'un même genre ne soit en réalité que des variétés (ce que l'on pourrait aujourd'hui assimiler à des souches), dont les caractères ont été influencés par des facteurs environnementaux, et dont l'apparition serait le fait d'un phénomène de génération spontanée à partir d'un ancêtre commun, et non pas de changements progressifs (Bowler 2003).

1.1.3.4. Lamarck, première théorie scientifique

Dans son ouvrage « *Philosophie Zoologique* » (de Lamarck 1809) Jean-Baptiste de Lamarck (1744-1829) présente la première théorie de l'évolution basée sur un raisonnement scientifique valable : le *transformisme*. Il propose que les caractères physiologiques acquis durant la vie d'un organisme, en réponse à son environnement, puissent être transmis à sa descendance. L'image d'Épinal de cet ouvrage est celle de l'animal au cou court et qui, de générations en générations, l'étire progressivement pour atteindre les feuilles en hauteur d'un arbre et donner ultimement naissance une nouvelle espèce (ici, la girafe en l'occurrence). Lamarck affine sa théorie en y exposant quatre lois dans « *Histoire Naturelle des Animaux Sans Vertèbres, vol I* » (de Lamarck 1815). Partisan de la génération spontanée, Lamarck voit l'évolution comme une complexification progressive du vivant, partant de l'organisme simple primordial vers ceux « plus évolués » par l'ajout successif de caractères acquis au fil des générations (*Première loi : La vie, par ses propres forces, tend continuellement à accroître le volume de tout corps qui la possède, et à étendre les dimensions de ses parties, jusqu'à un terme qu'elle amène elle-même*). L'environnement est le principal moteur évolutif, en générant des besoins auxquels l'individu répond par des adaptations sous forme de nouveaux caractères physiologiques (*Deuxième loi : La production d'un nouvel organe dans un corps animal, résulte d'un nouveau besoin survenu qui continue de se faire sentir, et d'un nouveau mouvement que ce besoin fait naître et entretient*). C'est l'utilisation ou la non-utilisation (*use and disuse*) des caractères qui dicte la

pérennité de ces adaptations (*Troisième loi: Le développement des organes et leur force d'action sont constamment en raison de l'emploi de ces organes*). Celles-ci qui, si elles sont maintenues, seront transmises à la descendance (*Quatrième loi : Tout ce qui a été acquis, tracé ou changé, dans l'organisation des individus, pendant le cours de leur vie, est conservé par la génération, et transmis aux nouveaux individus qui proviennent de ceux qui ont éprouvé ces changements*).

I.1.3.5. Darwin et la sélection naturelle

En 1859 dans « *On the Origins of Species* » (Darwin 1859), Charles Darwin (1809-1882) propose une vision de l'évolution différente, qui s'ancre autour de la sélection naturelle. Selon ce principe, les individus au sein d'une population ne sont pas identiques, mais présentent des variations plus ou moins importantes de leurs caractères phénotypiques. L'origine des variants sera un sujet longuement débattu à l'époque, mais Darwin pense qu'ils sont causés par l'environnement. Lorsque les individus sont soumis aux pressions de leur environnement, ceux avec des variations favorables, leur prodiguant une plus grande adaptation, auront une meilleure fitness (c'est à dire un succès reproducteur accru) et ces caractères favorables seront hérités par la descendance. *A contrario*, les individus présentant des caractères défavorables seront soit éliminés, soit auront une descendance réduite. Ce phénomène de sélection naturelle mène à l'augmentation et la disparition progressive des caractères favorables et défavorables, respectivement. Il évoque une seconde théorie, en lien avec la première, celle de la sélection sexuelle. Il s'agit d'un mécanisme selon lequel certains individus au sein d'une population auront une plus grande descendance que d'autres, de part leur meilleure capacité à acquérir des partenaires reproducteurs. Elle n'est pas associée à une lutte pour la survie de l'individu, mais à une compétition pour l'obtention du partenaire sexuel. Ce dernier est souvent sélectionné sur la base de caractères morphologiques ou comportementaux qui n'apportent pas forcément un avantage au niveau de la survie. Une fois de plus, ces caractères sexuels ne sont pas le fruit d'une adaptation de l'individu, mais bien d'une sélection au sein d'une diversité, présumée aléatoire, entre les individus d'une population. Cette théorie sera fortement détaillée dans son second ouvrage centré sur l'évolution, « *The Descent of Man, and Selection in Relation to Sex* » (Darwin 1872).

1.1.3.6. Les expériences de Weismann et la chute du transformisme

Une des grandes interrogations de l'époque, soulevée par les travaux de l'évolution, était d'identifier les mécanismes impliqués dans l'hérédité. C'est un aspect qui n'est peu ou pas développé par Lamarck, alors que Darwin proposera une explication dans son modèle de la pangénèse (Bowler 2003). Selon lui, des particules microscopiques nommées gemmules sont les porteurs des caractères héréditaires. On les retrouve dans tout l'organisme et elles auront tendance à s'accumuler dans les organes reproducteurs (Darwin 1892). Cette théorie a été largement critiquée, en particulier pour son aspect Lamarckiste, car Darwin n'excluait pas la possibilité que les caractères acquis durant l'existence des parents puissent être transmis à leur descendance de cette manière. Il est d'ailleurs intéressant de noter que, bien que les visions de l'évolution de Darwin et Lamarck aient été assez divergentes, Darwin n'a jamais totalement réfuté le transformisme. Il envisageait qu'il pouvait s'agir d'un mécanisme évolutif mineur, complémentaire à la sélection naturelle et sexuelle (Darwin 1892). La question reste toutefois très débattue dans la communauté scientifique. Francis Galton (1822-1911), cousin de Charles Darwin, effectue plusieurs expériences de transfusions sanguines entre des lapins de souches différentes. Il n'observe aucun impact sur la descendance, ce qui le mène à rejeter l'hypothèse de la pangénèse (Bowler 2003). Darwin conteste ces conclusions en expliquant qu'il n'a jamais écrit que les gemmules devaient transiter par le sang ou un fluide corporel (Darwin 1871).

La question de l'hérédité des caractères acquis reste débattue, et les travaux d'August Weismann (1834-1914) porteront un rude coup au transformisme. Cherchant à trancher une fois pour toute, il réalise une de ses plus célèbres expériences durant laquelle il coupe la queue de 68 souris sur 5 générations. Il n'observe aucune modification apparente des queues sur la progéniture : elles n'étaient ni absentes, ni raccourcies. Avec un regard contemporain, on pourrait simplement affirmer que la mutilation des queues de souris n'est pas un trait héréditaire (et certainement pas un phénotype adaptatif). À l'époque ces résultats menèrent à énoncer la théorie de la barrière de Weismann (*Germ Plasm Barrier Theory*) qui stipule que **(I)** les cellules germinales produisent des cellules somatiques et d'autres cellules germinales **(II)** les cellules germinales ne sont absolument pas affectées par ce que les cellules somatiques pourraient acquérir et expérimenter durant leur existence (Weismann *et al.* 1891). Ce concept sera par la suite une des pierres angulaires des théories de l'évolution à venir.

1.1.3.7. Émergence de la génétique

Dans les mêmes années, Gregor Mendel (1822-1884) effectue ses expériences sur l'hérédité de différents traits phénotypiques chez les petits pois. Il introduit les concepts de caractères récessifs et dominants (Mendel 1866). Ses travaux sur l'hérédité ne marqueront pas ses contemporains, mais seront redécouverts par Hugo De Vries (1848-1935, à qui on doit le terme « mutation ») et Carl Correns (1864-1933) au début du XX^{ème} siècle (Bowler 2003). Fin 1890 – début 1900, le botaniste Danois Wilhelm Johannsen (1857-1927) observe dans des lignées pures de pois (*i.e.* sans variabilité, et probablement homozygotes, sur les traits observés) que la taille des cosses de pois suit une distribution normale (Johannsen 1903). C'est un des premiers exemples de plasticité phénotypique, qui le mène à proposer les notions de « génotype » et « phénotype » (Johannsen & Selskab 1903; Johannsen 1905). Il est aussi crédité de la première définition du gène : *« le terme gène est dénué de toute hypothèse; il exprime seulement l'évidence que plusieurs caractéristiques de l'organisme sont spécifiées dans les cellules germinales par des conditions spéciales, fondations, et déterminants qui sont présents de manières uniques, séparées et donc indépendantes – en bref, précisément ce que nous souhaitons appeler gènes »* (Johannsen 1909). Ces avancées seront marquées par le conflit entre les partisans des théories Mendéliennes dont la figure de proue est William Bateson (1861-1926), à qui on doit l'appellation « génétique », et le courant « biostatisticien » dont les hérauts sont Karl Pearson (1857-1936) et Walter Frank Raphael Weldon (1860-1906). Ces derniers basent leurs travaux sur la mesure quantitative de la variation de traits phénotypiques au sein de populations et affirment que le modèle de la génétique mendélienne est insuffisant pour expliquer toute la diversité observée dans les caractères (Bowler 2003).

Une avancée substantielle est l'œuvre de Thomas Hunt Morgan (1866-1945) qui, grâce à ses expériences de croisements chez *Drosophila melanogaster*, identifie les chromosomes comme étant les porteurs de l'information héréditable (Morgan 1915). En parallèle, les travaux de Ronald Fisher (1890-1962) posent les bases de la génétique des populations et font le lien entre les courants mendélien et biostatisticien. Il met en évidence le fait que la variation des caractères s'explique par l'interaction de plusieurs gènes. L'environnement impose une pression de sélection sur certains caractères, menant à changer la fréquence d'apparition de ces gènes dans une population (fréquence allélique) (Fisher 1930). Hermann Joseph Muller (1890-1967) conclura que les mutations pouvant être

à l'origine des variations de caractères surviennent aléatoirement, et ne sont pas ciblées. D'autres scientifiques, tels que Theodosius Dobzhansky (1900-1975) et Edmund Brisco « Henry » Ford (1901-1988) contribueront à la progression de la discipline en montrant que la diversité génétique au sein des populations naturelles est plus élevée que ce qui était préconçu, et que l'impact de l'environnement joue un rôle de premier plan dans le maintien de cette diversité (Bowler 2003). À la même période Richard Woltereck (1877-1944) introduit l'idée de norme de réaction pour expliquer la diversité des phénotypes produits par un seul génotype, en fonction des changements de l'environnement.

1.1.3.8. Théorie synthétique de l'évolution

Ces avancées sonnent le glas des dernières écoles de pensée Lamarckistes. Leurs fondements sont insolubles dans les mécanismes de plus en plus tangibles de la génétique (Bowler 2003). La seule exception notable à cette tendance se situe en ex-Union Soviétique. Entre la fin des années 1920 et 1960, un agronome du nom de Trofim Lysenko (1898 – 1976) rejette complètement la génétique et l'hérédité mendélienne. Il soutient que les traits héréditaires sont le résultat de la concentration de facteurs environnementaux divers, assimilés par les organismes au fil des générations précédentes. Il se sert de cette théorie pour promouvoir une technique de vernalisation (traitement des semences à forte humidité et faible température) soi-disant révolutionnaire (pouvant jusqu'à quadrupler les récoltes sur plusieurs générations). La vernalisation a bien un effet sur quelques variétés, mais sa généralisation à toutes les cultures était inadaptée. Dans les faits, les idées de Lysenko tenaient beaucoup plus de la politique de propagande soviétique que de la véritable science, et son œuvre (et surtout ses conséquences désastreuses sur la famine) contribuera par la suite à un rejet encore plus marqué des théories Lamarckistes (Gordin 2011). Sélection naturelle, génétique mendélienne, hérédité des chromosomes et génétique des populations (entres autres) sont réunis par Julian Huxley (1887-1975) dans sa théorie synthétique de l'évolution (Huxley 1943). Brièvement, cette théorie définit les processus d'évolution et d'adaptation spécifiques à un environnement comme des processus lents, au travers de nombreuses générations, dirigés par des mutations aléatoires et la sélection des individus porteurs des mutations prodiguant la meilleure fitness (Huxley 1943; Smocovitis 1992). L'ouvrage d'Huxley a un impact important sur ses contemporains et impose la sélection naturelle comme moteur principal de la dynamique évolutive (Bowler 2003).

I.1.3.9. Progrès moléculaires et génocentrisme

Au milieu du XX^{ème} siècle, la découverte de la structure chimique de l'ADN par Rosalind Franklin (1920-1958), Francis Crick (1916-2004) et James Watson (né en 1928) (Watson & Crick 1953), ainsi que celle du code génétique par Robert Holley (1922-1993), Marshall Warren Nirenberg (1927-2010) et Har Gobind Khorana (1922-2011) sont des étapes charnières dans la compréhension des mécanismes de l'hérédité (Bowler 2003). Les premières hypothèses de l'horloge moléculaire par Linus Pauling (1901-1994) et Émile Zuckerkandl (1922-2013), en comparant des séquences de protéines homologues, ouvrent la porte à une estimation du temps de spéciation entre deux organismes (Zuckerkandl & Pauling 1962). À la fin des années 60, Motoo Kimura (1924-1994) reprend cette idée dans sa théorie neutraliste de l'évolution, stipulant que la plupart des mutations n'ont aucun impact sur la fitness d'un organisme, et que la dérive est une force évolutive plus puissante que la sélection naturelle (Dietrich 1994). L'ADN et son unité d'héritabilité, le gène, deviennent rapidement les points centraux des mécanismes évolutifs et de la sélection naturelle. Cette vision sera confortée par la découverte chez les procaryotes du transfert horizontal de gènes en 1961 (Watanabe & Fukasawa 1961) et culminera avec la publication du « gène égoïste » par Richard Dawkins (né en 1941) en 1976 (Dawkins 2006). Cet ouvrage présente l'individu (peu importe son espèce) comme un « *véhicule robotisé, programmé aveuglément pour préserver des molécules égoïstes nommées gènes* » (Dawkins 2006).

I.1.3.10. Importance des interactions

La seconde moitié du XX^{ème} siècle s'inscrit dans l'étude du rôle des interactions intra et interspécifiques au sein de l'évolution. Dans sa théorie de la Reine Rouge, Leigh Van Valen (1935-2010) pose les bases de la coévolution, décrivant les interactions entre hôtes et pathogènes comme une perpétuelle course aux armements sans vainqueur, mais générant une force évolutive importante (Van Valen 1973). Les rôles dans l'évolution de la reproduction sexuée et de la sélection sexuelle sont débattus. Elles sont initialement perçues dans la continuité des idées de Van Valen comme des mécanismes générant de la diversité pour répondre rapidement à la pression des pathogènes (Ebert & Hamilton 1996). D'autres y voient plutôt un filtre évolutif permettant d'éliminer les changements et aberrations génétiques majeurs, et permissif seulement aux polymorphismes ayant un

impact neutre ou faible (MacIntyre & Clegg 2013). La question du comportement altruiste chez certaines espèces (comme les insectes eusociaux) est également soulevée par William Donald Hamilton (1936-2000). Il décrit comment une interaction comportementale altruiste peut être favorisée par la sélection naturelle en fonction du degré de parentalité entre les individus, et si les risques (du comportement altruiste) sont moindres que les bénéfices (Hamilton 1964). Cela s'explique par le fait que plus deux individus sont apparentés (ex : parents/enfants ou frères/sœurs), plus ils partagent un patrimoine génétique commun. Il est donc avantageux de favoriser la fitness de l'autre car elle favorise la transmission d'une partie des gènes en commun (et ce, même au détriment de sa propre fitness, tant que les bénéfices surpassent le coût). C'est ce qu'Hamilton nomme la sélection de parentèle (*kin selection* en anglais) (Hamilton 1964), une théorie qui partage plusieurs points communs avec le concept de gène égoïste de Richard Dawkins.

Les travaux d'Hamilton seront remarqués par George Price (1922-1975). Initialement chimiste sans formation en biologie, il se base sur la théorie de la sélection des parentèles pour énoncer son équation éponyme, modélisant les changements de fréquences alléliques dans une population (Price 1970). L'histoire personnelle de George Price (que m'a fait découvrir Raymond Pierce) est à la fois fascinante et tragique, et mérite que l'on ouvre une parenthèse pour la raconter. Peu de temps après sa rencontre avec Hamilton, Price, pourtant athée et ultra-rationaliste, trouve qu'il y a trop de coïncidences dans sa vie pour qu'elles n'aient pas une origine divine. Il se convertit au Catholicisme et Jésus lui apparaît peu de temps après en vision, l'incitant à transmettre son équation et ses fondements à l'humanité (Harman 2010). Il devient cependant de plus en plus déprimé par les implications de son équation, *i.e.* que bonté et comportements altruistes ne sont pas véritablement désintéressés, puisqu'il s'agit d'une adaptation pour promouvoir son propre patrimoine génétique. Dans l'espoir de valider ou d'invalidier sa théorie, Price effectue de plus en plus d'actes de générosité aléatoires envers des inconnus. Il héberge des sans-abris et distribue presque toutes ses possessions à des nécessiteux, au point de se retrouver lui-même sans domicile fixe. Incapable de conclure sur sa théorie, il se suicide le 6 janvier 1975 ; son corps sera identifié par son ami Hamilton (Harman 2010). Price apporte deux autres contributions majeures à la biologie évolutive : une nouvelle interprétation du théorème fondamental de la sélection naturelle de Ronald Fisher (Price 1972) et le concept de stratégie évolutive stable avec John Maynard Smith (1920-2004) (Maynard Smith & Price 1973).

I.1.4. Hérité des caractères acquis

I.1.4.1. Victor Jollos et les *Dauermodifikationen*

De l'ensemble de ces travaux naquit, durant la deuxième moitié du XX^{ème} siècle, une doctrine présentant l'ADN comme l'unique « plan de construction du vivant ». Les séquences nucléotidiques sont interprétées au travers du code génétique et sont



Figure 4 : Victor Jollos en 1934. Photographie d'archive du Marine Biological Laboratory, <http://hpsrepository.mbl.edu>.

globalement perçues comme étant les seuls porteurs de l'information héritable. Le phénotype est alors expliqué comme étant le résultat de l'expression de gènes sous l'influence de l'environnement. Le matériel génétique permet une certaine flexibilité à l'organisme pour répondre aux conditions environnementales fluctuantes sous la forme de la plasticité phénotypique. Pourtant, il y a maintenant presque un siècle, un scientifique allemand du nom de Victor Jollos (**Figure 4**) a mis en évidence que certains phénotypes pouvaient être induit par des conditions environnementales particulières, et étaient en mesure de persister durant quelques générations en l'absence du stimulus initial, avant de disparaître progressivement. Ses travaux portaient sur *Paramecium spp* et *D. melanogaster* (Jollos 1921; 1924; 1934; Brink 1941). Il a baptisé ces phénomènes

« *Dauermodifikationen* » (littéralement : modifications de longue durée), mais ils sont aussi aujourd'hui connus sous les termes de « résonance phénotypique », « hérédité non mendélienne », « phénotype transgénérationnel », ou encore « *soft inheritance* ». Chez *Paramecium*, son modèle de choix pendant presque 20 ans, il utilise des lignées issues de reproduction asexuée ou parthénogénique et les expose à divers stress environnementaux, tels que des chocs thermiques ou des métaux lourds, et observe l'apparition de nouvelles adaptations en réponses à ces stress. Il constate une semi-hérédité, selon ses termes, car

les adaptations sont transmises à la descendance pendant plusieurs centaines de générations, mais ne sont pas incorporées dans l'hérédité permanente de la souche et finissent par disparaître en l'absence du stress (Jollos 1924). Il travaille ensuite sur l'effet de chocs thermiques chez les drosophiles, et remarque l'apparition de certains traits perdurant sur quelques générations (Jollos 1934). Les idées de Jollos n'ont pas rencontré un accueil très chaleureux au sein des cercles évolutionnistes de son époque. Plusieurs facteurs ont joué un rôle dans l'absence de reconnaissance de ses travaux par ses contemporains. Si au début de sa carrière, Jollos bénéficie d'un statut renommé en Allemagne, il finit destitué de son poste de généticien à l'Institut Kaiser Wilhelm de Berlin en 1933 à cause des politiques antisémites de l'époque. Contraint à l'exil, il s'installe à l'Université du Wisconsin (Madison, USA) où sa réputation reste à faire. Il peine à trouver sa place dans les thématiques de l'université, ainsi qu'à obtenir des financements (Dietrich 1996). Son arrivée aux États-Unis coïncide avec son changement de modèle biologique pour la drosophile. Ses résultats gagnent alors peu de crédibilité aux yeux des autres académiciens, car le modèle d'hérédité que Jollos préconise va à l'encontre des courants de pensées de Thomas Hunt Morgan, qui jouit alors d'une très forte influence (Dietrich 1996). De plus, plusieurs chercheurs tentent de reproduire, sans succès ses résultats (Brink 1941; Dietrich 1996). Après son décès prématuré en 1941, les *Dauermodifikationen* sont tombées dans l'oubli et ont été majoritairement considérées comme des artéfacts expérimentaux.

1.1.4.2. Exemples d'hérédité non-mendélienne

Toutefois, principalement depuis le début du XXI^{ème} siècle, de plus en plus de publications font état de différents organismes, provenant de phyla variés, développant des adaptations héréditaires et réversibles en réponse à des changements environnementaux, en l'espace de quelques générations seulement (Agrawal *et al.* 1999; Morgan & Whitelaw 2008; Youngson & Whitelaw 2008; Jensen 2013).

Chez les végétaux, ces phénomènes commencent à être bien documentés et on retrouve des exemples chez plusieurs espèces en réponse à une grande variété de stimuli environnementaux. L'herbivorie est un inducteur important de défense. En réponse à la chenille spécialiste *Pieris rapae*, le radis sauvage *Rafus raphanistrum* augmente de 30% sa densité de poils trichomes et de dix fois ses concentrations cellulaires d'indoles

glucosinolates, une substance toxique pour l'insecte (Agrawal *et al.* 1999). Ces deux traits réduisent le succès de broutage par la chenille, et les graines issues des plantes induites auront naturellement ces défenses, même sans avoir rencontré directement le même stress. D'autres plantes, comme *Mimulus guttatus*, *Lotus wrangelianus* (Holeski *et al.* 2012) et *Arabidopsis thaliana* (Rasman *et al.* 2012) induisent similairement des défenses transgénérationnelles en réponse à des insectes brouteurs. Chez le tabac *Nicotiana glauca*, l'infection de pieds parentaux par le virus de la mosaïque du tabac (TMV) induit une résistance chez la descendance qui perdure pour quelques générations (Roberts 1983). Plus d'une vingtaine d'autres exemples impliquant des inducteurs chimiques ou des variations climatiques ont été publiés depuis une quinzaine d'années (Holeski *et al.* 2012).

Des cas s'apparentant à des *Dauermodifikationen* sont documentés chez les métazoaires. Un des plus emblématiques est certainement le phénomène de cyclomorphose observée chez plusieurs espèces de crustacés du genre *Daphnia*. En réponse à l'apparition d'un prédateur dans leur entourage, les *Daphnia* développent un casque, un aiguillon, une collerette dentée ou une augmentation de taille (**Figure 5**). Ces adaptations morphologiques réduisent le succès de capture du prédateur (Laforsch & Tollrian 2004). La descendance hérite de ces caractères, même si elle n'a jamais été confrontée directement au prédateur. En l'absence du prédateur, ces traits sont maintenus durant un petit nombre de générations (3-5 habituellement) avant de régresser progressivement vers le phénotype « normal ». La vitesse d'apparition de ce caractère adaptatif héritable au sein de la population de *Daphnia* ne peut être expliquée par le processus classique de mutation/sélection.

Chez les vertébrés, un exemple récent est celui du poisson clown *Amphiprion melanopus*. Ce poisson a développé une résistance héritable à l'accroissement de la température et à l'augmentation de l'acidité de l'eau beaucoup plus rapidement que ne pourrait l'expliquer la sélection naturelle (Miller *et al.* 2012). Une série d'étude sur des cas de restriction alimentaire chez l'être humain en Suède a montré un impact sur le risque de maladie cardiovasculaire, de diabète et sur la croissance, perdurant sur au moins trois générations (Kaati *et al.* 2002; 2007; Bygren *et al.* 2014). D'autres cas de réponses transgénérationnelles chez l'être humain ont été répertoriées (Guerrero-Bosagna & Skinner 2011; Pembrey *et al.* 2013). On retrouve également des effets de l'environnement induisant

des réponses comportementales. Entre autres, le stress engendre des changements héréditaires de stratégie alimentaire et de dominance sur le coq *Gallus gallus* (Nätt 2008). Un des exemples les plus cités est celui de l'effet du soin maternel chez les souris. Les souriceaux séparés de leur mère et n'ayant pas eu accès aux soins sont beaucoup plus susceptibles d'avoir des comportements anormaux. À leur tour, ces souris ne prodigueront pas de soins à leur progéniture, ce qui cause le maintien du phénotype d'une génération à une autre (Weaver *et al.* 2004).

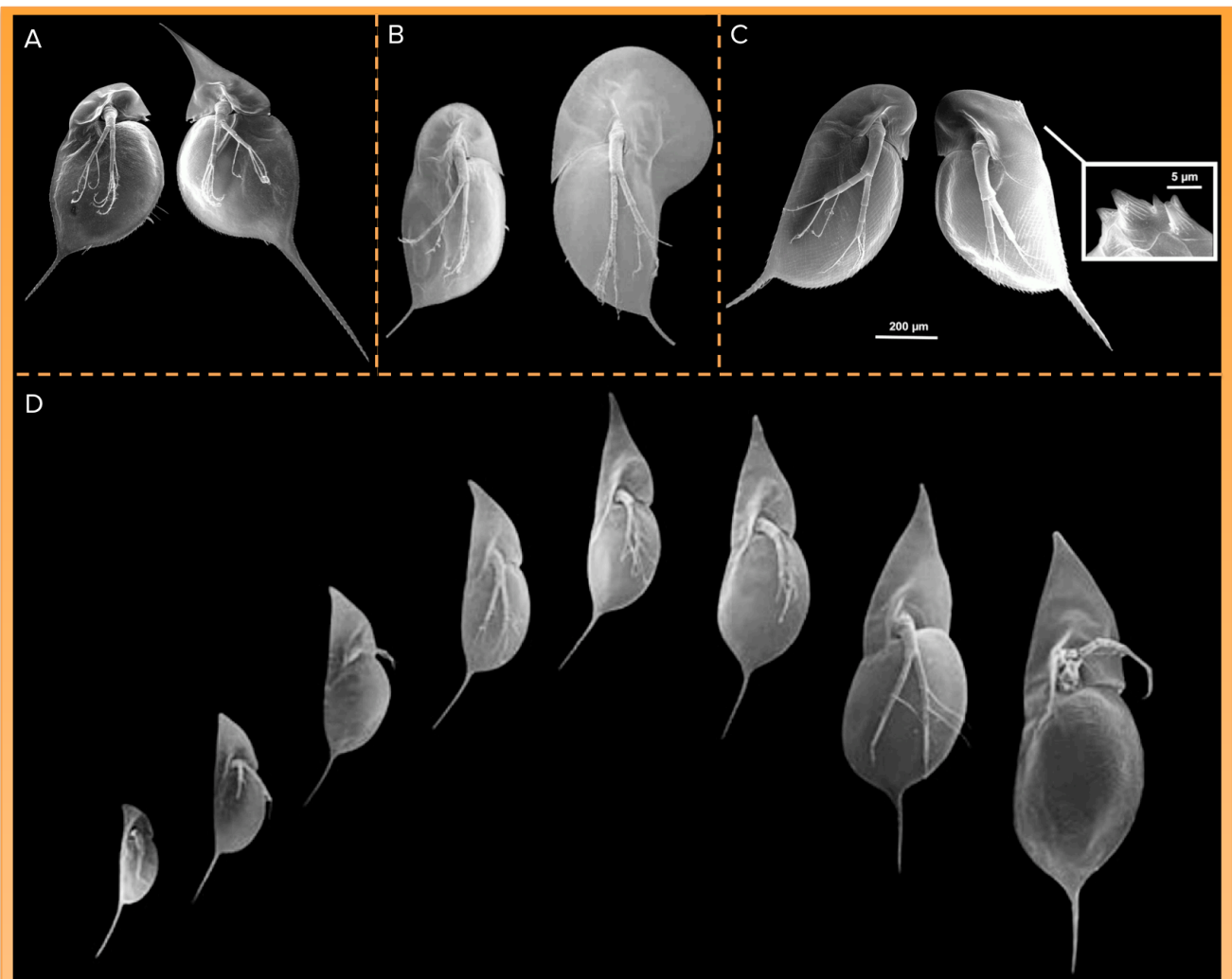


Figure 5 : Exemples de *Dauermodifikationen* chez le genre *Daphnia*. À gauche, le phénotype normal, à droite, le phénotype de réponse à la présence d'un prédateur. **(A)** *Daphnia lumholtzi* avec et sans casques et aiguillon caudal. **(B)** *Daphnia longicephala* avec et sans casque. **(C)** *Daphnia pullex* avec et sans collier denté. **(D)** *Daphnia cucullata* en l'absence (extrême gauche) et présence (extrême droite) de prédateurs, ainsi que toutes les formes intermédiaires. Figures tirées de Laforsch & Tollrian (2004).

Ne suivant pas une logique d'hérédité mendélienne, ces observations sont susceptibles de remettre en question certains aspects de la théorie synthétique de l'évolution, en particulier la barrière de Weismann. Cela soulève deux grandes questions :

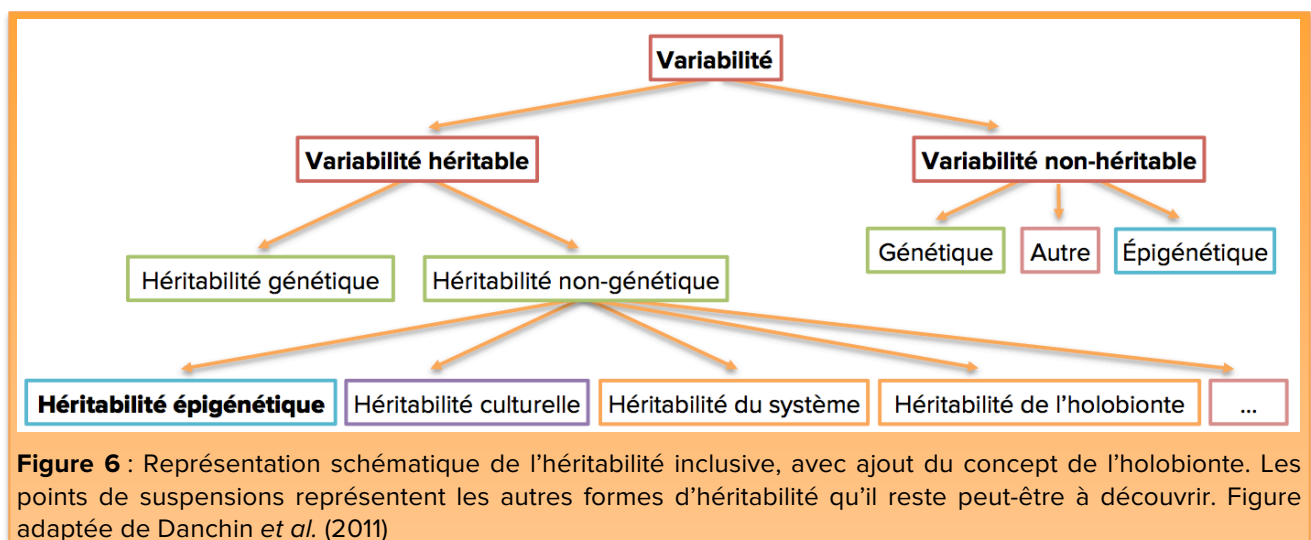
ces *Dauermodifikationen* sont-elles des phénomènes isolés, et limités à quelques organismes, ou répandus au sein du vivant (et quel est leur importance dans l'évolution), et quels en sont les mécanismes moléculaires sous-jacents?

I.1.4.3. Origines possibles des *Dauermodifikationen*

Les contemporains de Jollos ont vu en ses travaux sur *Paramecium* la preuve d'une hérédité cytoplasmique, au moins aussi importante que l'hérédité nucléaire (génétique). Toutefois, Jollos était dubitatif vis à vis de cette idée, du fait que dans sa conception, les facteurs cytoplasmiques induit par l'environnement devaient forcément se diluer à chaque division cellulaire. Ainsi, les phénotypes n'auraient pas pu se maintenir sur plus de quelques générations (alors qu'il observait un maintien sur plusieurs centaines) (Dietrich 1996). Dans le cas de la cyclomorphose chez *Daphnia*, des kairomones émises par les prédateurs, ou des signaux chimiques provenant d'autres proies victimes des mêmes prédateurs, induisent le phénotype de défense (Weber & Declerck 1997; Stabell *et al.* 2003). La persistance de ces médiateurs chimiques dans l'environnement expliquerait alors le maintien du phénotype même après la disparition du prédateur. La paramutation, un mécanisme par lequel un allèle peut induire une modification héritable sur un autre allèle, à un locus donné, un phénomène bien connu chez le maïs (Alleman *et al.* 2006), peut expliquer certains phénotypes d'hérédité non-mendélienne (Goldberg *et al.* 2007). On peut aussi penser dans une certaine mesure au phénomène d'épistasie. Ainsi, l'expression de certains gènes (et donc d'un phénotype) est dépendante de la présence d'un ou plusieurs autres gènes sur d'autres loci. Si l'hérédité des gènes suit une logique mendélienne, ce ne sera pas forcément le cas du phénotype.

Les *Dauermodifikationen* semblent être en contradiction avec les dogmes classiques de la génétique mendélienne et de la théorie synthétique de l'évolution. S'il a été longtemps admis que seule l'information contenue dans la séquence génétique était responsable des variations phénotypiques, il est de plus en plus clair que ce n'est pas le cas. De plus en plus d'exemples mettent en avant des changements phénotypiques entre individus d'une même espèce, changements que les polymorphismes génétiques ne suffisent pas à expliquer (Maher 2008; Slatkin 2009; Manolio *et al.* 2009; Zuk *et al.* 2012). C'est ce que l'on appelle l'hérédité manquante (*missing heritability*). Par exemple, chez

l'être humain, les études d'association à l'échelle du génome (*genome wide association studies*) ne permettent d'expliquer par des changements génétiques qu'en moyenne 20-30% de la variation de traits étudiés (Zuk et al. 2012). En 1990, John Maynard-Smith développe le modèle de la double hérédité (*dual inheritance system*) dans lequel il propose deux formes d'hérédité. La première est basée sur les séquences d'ADN, et la seconde sur des mécanismes permettant une réponse héritable et pourtant flexible aux fluctuations environnementales : le code épigénétique (Maynard Smith 1990). Cette théorie a été complétée et raffinée pour y inclure d'autres types d'héritabilités influençant les phénotypes : comportementale (culturelle ou sociale), environnementale et parentale (Jablonka et al. 1992; Jablonka & Lamb 2005; 2008; Danchin et al. 2011; Danchin 2013). Étienne Danchin propose une synthèse des facteurs héritables et non-héritables générant de la variabilité dans son concept d'héritabilité inclusive (*inclusive heritability*, Danchin et al. 2011; Danchin 2013). Cette vision, à laquelle j'ai rajouté l'héritabilité de l'holobionte (à savoir l'organisme et ses communautés de micro-organismes) est représentée sur la **Figure 6**. De plus en plus de travaux ont mis en évidence que les communautés microbiennes (bactériennes par exemple, mais pas uniquement) avec lesquelles coexiste un organisme peuvent avoir un impact direct sur certains phénotypes et sur la fitness, et même parfois être directement transmises à la descendance (Dunbar et al. 2007; Rosenberg et al. 2007; Zilber-Rosenberg & Rosenberg 2008; Rosenberg et al. 2009; Sharon et al. 2010). Au niveau moléculaire, les mécanismes épigénétiques représentent donc une alternative de choix pour expliquer une proportion de cette hérédité manquante, surtout chez les organismes sans hérédité sociale et culturelle, et sont de bons candidats pour la compréhension des *Dauermodifikationen*.

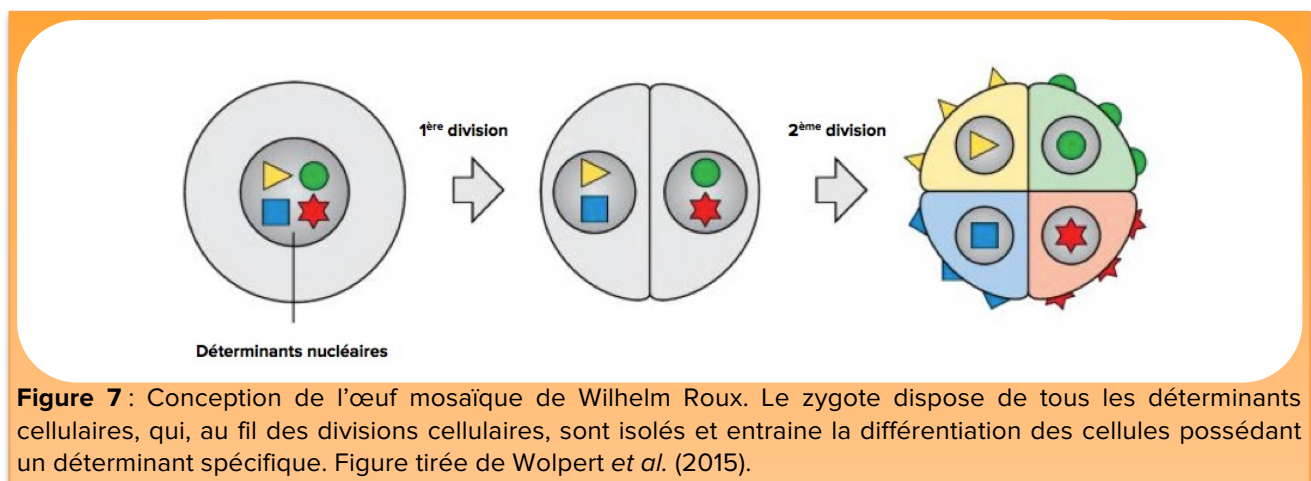


I.2. L'ÉPIGÉNÉTIQUE

I.2.1. Bref historique de la discipline

I.2.1.1. Embryologie, épigénèse et préformationnisme

Les origines de l'épigénétique moderne remontent à la fin du XIX^{ème} siècle, et trouvent leurs sources dans l'étude du développement embryonnaire. Deux écoles de pensées s'affrontent alors sur les mécanismes pouvant expliquer les processus de transformation menant un zygote unicellulaire à un organisme complet. Les préformationnistes voient les organes déjà tous établis en version miniature dans le zygote, et grandissant au fil du développement. De l'autre côté, les partisans de l'épigénèse proposent que le processus de développement incorpore des réactions chimiques incluant des constituants solubles, et se définit par un plan d'organisation complexe (Felsenfeld 2014). La question n'est pas nouvelle, et se pose déjà dans l'Antiquité. Aristote (proche épigénèse) et Platon (proche préformationniste) opposait déjà ces deux conceptions (Hurd 2011; Rankin 2013). La théorie préformationniste sera prédominante jusqu'au XIX^{ème} siècle. Entre temps, William Harvey (1578-1657) pense que les deux modes de développement coexistent, mais chez des organismes différents : la préformation se limite aux organismes « inférieurs » (ex : les insectes) alors que l'épigénèse est un processus plus noble réservé aux espèces plus complexes (*i.e.* les vertébrés) (Hall 2006). Seule une minorité de naturalistes, tels que Pierre Louis Moreau de Maupertuis (1698-1759) et Georges-Louis Leclerc (comte de Buffon) seront partisans notoires de l'épigénèse (Roger 1997). Au XIX^{ème} siècle, Wilhelm Roux (1850-1924) avance la théorie de l'œuf mosaïque (**Figure 7**). Il propose par cela que tous les constituants sont préformés dans le zygote et se séparent lors des divisions cellulaires (Counce *et al.* 2012).



La théorie de Roux est en contradiction avec les expériences de Hans Driesch (1867-1941) et Oscar Hertwig (1849-1922) sur les blastomères d'oursins qui continuent à se développer même s'ils sont séparés en morceaux (Gilbert 2000). Le concept d'épigénétique se développe grâce à une volonté de relier développement embryonnaire et génétique. Thomas Hunt Morgan est un des premiers à soulever un *a priori* paradoxe entre ontologie et hérédité dans sa Théorie du Gène (Morgan 1917) où il questionne : « *Si les caractères de l'individu sont déterminés par les gènes, pourquoi toutes les cellules d'un organisme ne sont-elles pas identiques ?* ». Edmund Beecher Wilson (1878-1939) critique la théorie de l'œuf mosaïque car elle n'est pas en accord avec l'hérédité chromosomique (Wilson 1893). Le préformationnisme disparaît définitivement avec la découverte de l'induction embryonnaire (signal biochimique initiant la différenciation des tissus) par Hans Spemann (1869-1941) et Hilde Mangold (1898-1924) (Spemann & Mangold 2001). Les mécanismes derrière l'épigénèse restent néanmoins encore vagues et mal compris.

1.2.1.2. Le paysage épigénétique de Waddington

Au début des années 1930, Conrad Hal Waddington (1905-1975) s'intéresse au contrôle génétique du développement. Il élabore des diagrammes décrivant les chemins développementaux alternatifs survenant en fonction des déterminants moléculaires. Il présume que ces déterminants agissent sur des gènes spécifiques, générant un phénotype (Waddington 1939; 1940). En 1942, il publie son article le plus cité, « l'épigénotype » (Waddington 1942b) dans lequel il fait la première mention du terme « épigénétique ». Il s'agit à ce moment là d'une définition assez vaste incluant « *tous les évènements menant à la mise en place du programme génétique du développement* ». L'accent est mis sur le rapport entre génotype et phénotype, pour voir comment ils s'articulent. Entre les deux se trouvent un ensemble complexe de processus environnementaux, l'épigénotype. Selon lui, il faut mettre l'emphasis sur les mécanismes causaux associant génotype et phénotype. En 1957 dans « la stratégie des gènes » (Waddington 1957), il formule le concept du paysage épigénétique, une métaphore de la manière dont la génétique peut contrôler et influencer le développement. Dans une de ses représentations les plus connues (**Figure 8**), il compare le destin cellulaire à une bille, qui glisse sur un terrain dont le relief n'est pas prédéfini, mais modulée par les gènes et leurs mutations. Les « montagnes » séparant les vallées dans lesquelles les billes voyagent représentent l'irréversibilité de la différenciation cellulaire.

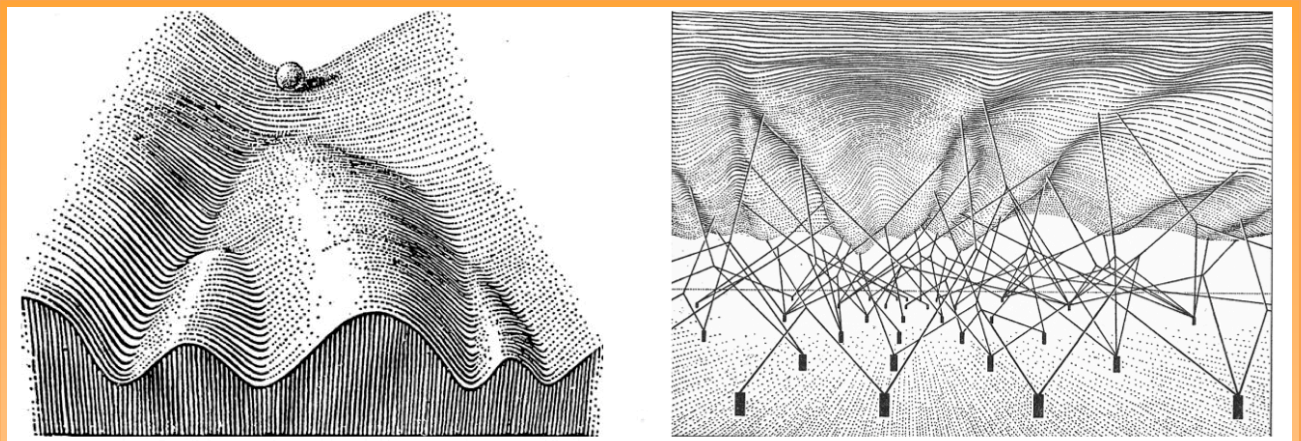
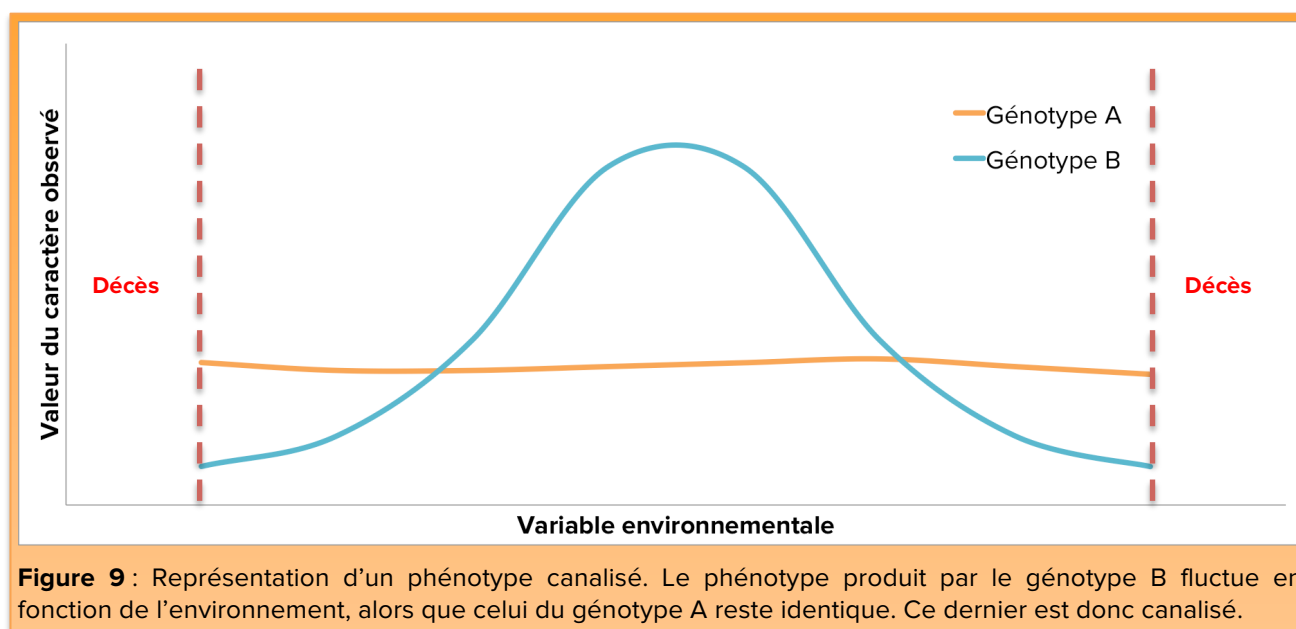


Figure 8 : Paysage épigénétique de Waddington. À gauche, la bille représentant le destin cellulaire glisse sur le paysage épigénétique qui est façonné par l'interaction des gènes et de l'environnement, représentés à droite. Figure tirée de Waddington (1957).

Waddington fait un parallèle évolutif, montrant que la mutation, par son action sur le développement et la morphologie, joue un rôle important dans l'évolution et la sélection naturelle (van Speybroeck *et al.* 2002). On lui doit aussi l'assimilation génétique, avec deux exemples chez *D. melanogaster*. Il observe que des chocs thermiques sur des pupes de drosophiles engendrent un phénotype « *crossveinless* » (absence de veines transversales dans les ailes) chez les mouches adultes (Waddington 1953). Après répétitions du choc thermique sur plusieurs générations, les drosophiles finissent par présenter ce caractère *crossveinless* même en l'absence du stress environnemental. De la même manière, en effectuant des expériences d'exposition de drosophiles à des vapeurs d'éther, Waddington remarque que celles-ci se développent avec deux thorax (Waddington 1956). Il sélectionne durant une vingtaine de génération les drosophiles présentant ce phénotype, jusqu'à obtenir des mouches présentant naturellement le double thorax, même sans induction aux vapeurs d'éther. Il conclut que la sélection (naturelle ou artificielle) favorise un chemin spécifique sur le paysage épigénétique et incorpore ce phénotype dans l'hérédité génétique (Waddington 1953; 1956). L'assimilation génétique est complémentaire de la canalisation, un autre concept qu'il développe dans le même article. Il s'agit d'une mesure de la capacité d'une population à produire un phénotype identique peu importe la variabilité de l'environnement ou des génotypes au sein de la population (Waddington 1942a; 1953) (**Figure 9**). Le rôle de la canalisation est de produire un phénotype précis, peu importe les fluctuations mineures de l'environnement. Pour certains contemporains de Waddington, ses résultats sur l'assimilation génétique sont un retour aux sources de l'hérédité des caractères acquis et du lamarckisme.



1.2.1.2. L'épigénétique moderne

S'il avait été montré depuis les expériences de Thomas Hunt Morgan que les chromosomes étaient hérités par toutes les cellules lors du développement, il n'y avait aucune preuve que tous les gènes portés par les chromosomes se trouvaient bien à l'identique dans chaque type cellulaire (Felsenfeld 2014). Cette démonstration sera faite par Ron Laskey (né en 1945) et John Gurdon (né en 1933) qui observent que la transplantation du noyau d'une cellule somatique adulte dans un œuf énucléé permet une embryogénèse directe (Laskey & Gurdon 1970). La découverte du rôle de la méthylation de l'ADN dans l'inactivation du chromosome X offre un premier mécanisme de contrôle épigénétique (Riggs 1975). La méthylation des cytosines devient un centre d'attention dans les années qui suivirent, jusqu'à aujourd'hui, en particulier pour son rôle dans la régulation de l'expression des gènes (Felsenfeld 2014). La structure de la chromatine, c'est à dire l'ADN nucléaire et les protéines interagissant intimement avec, devient également un sujet d'étude. L'ADN nucléaire est enroulé autour de nucléosomes, des octamères d'histones, et Ellen et Edgar Stedman proposent en 1950 (Stedman 1950) que ces dernières puissent être impliqués dans l'expression génique. Cette idée fera son chemin jusqu'à l'hypothèse du code des histones par Brian Turner vers la fin des années 1990 (Turner 1993; 2002). Au XXI^{ème} siècle, l'essor des technologies de séquençage massif permet de découvrir le rôle des petits ARNs non-codants et leur implication dans les phénomènes épigénétiques. Il devient plus tard évident que ces différentes marques épigénétiques et les mécanismes initiant leur lecture ou écriture sont intimement reliés (Felsenfeld 2014). D'autres possibles

porteurs d'information épigénétiques seront discutés plus loin dans ce manuscrit. Au fur et à mesure des découvertes des sources d'information épigénétique, la définition de la discipline se transforme progressivement. Robin Holliday la qualifie en 1990 « *d'étude des mécanismes de contrôles temporels et spatiaux de l'activité génique durant le développement des organismes complexes* » (Jablonka & Lamb 2002). Cette définition reste proche des origines embryologistes de Waddington, et il la fera évoluer plus tard en « *l'étude des changements de l'expression génique, survenant dans des organismes avec des cellules différenciées, et l'héritabilité mitotique des profils donnés d'expression génique* » (Jablonka & Lamb 2002). Elle sera étendue par Vincenzo Russo et Arthur Riggs qui lui donne une plus grande ouverture en la consacrant comme « *l'étude des modifications de la fonction des gènes, héritable mitotiquement ou méiotiquement, et qui ne peuvent être expliquées par des modifications de la séquence d'ADN* » (Russo et al. 1996). Cette version ne convient cependant pas à Andrew Bird, car au début des années 2000, il est évident que tous les processus épigénétiques ne sont pas forcément héréditaires. Il tente de redéfinir la discipline en parlant « *d'adaptations structurelles de régions chromosomiques de manière à inscrire, signaler ou perpétuer des états d'activité altérés* » (Bird 2007). Plus récemment, le consortium du Roadmap Epigenomics Project lui donne comme définition « *les modifications héréditaires dans l'activité génique et l'expression, mais également les altérations stables et de longue durée dans le potentiel transcriptionnel qui ne sont pas nécessairement héréditaires* » (<http://www.roadmapepigenomics.org/overview>). Je complète par trois définitions proposées par Francesca Merlin (CR2, Institut d'Histoire et de Philosophie des Sciences et des Techniques, Paris) lors du l'école thématique Épigénétique en Écologie et Évolution (18-22 mai 2015, Saint Pierre d'Oléron). Ces définitions vont d'une plus générale à une plus pointue et moléculaire.

I) Biologie Évolutive du Développement (Évo-Dévo) Épigénétique : étude des mécanismes développementaux à l'origine du phénotype et de sa modification au cours de l'évolution (sans altérer la séquence d'ADN). Cette définition est associée à la génétique du développement et à la biologie évolutionnaire du développement.

II) Waddington Épigénétique : étude des mécanismes causaux en œuvre durant le développement, menant du génotype au phénotype. Cette définition est associée à la biologie du développement.

III) Riggs & Holliday Épigénétique : étude des changements de la fonction des gènes mitotiquement et/ou méiotiquement héritablement qui ne peuvent être expliqués par des changements de la séquence d'ADN. Cette définition est associée à la génétique et/ou l'épigénétique moléculaire et est inspirée par Riggs (1975) et Holliday & Pugh (1975).

On ne peut pas contester le lien entre épigénétique et développement, mais j'ai une vision plus moléculaire et la Riggs & Holliday épigénétique est ma définition de prédilection.

1.2.2. Porteurs d'information épigénétique

Les définitions actuelles de l'épigénétique restent toutefois relativement larges, et sont utilisées pour décrire des phénomènes dont les bases moléculaires et les mécanismes sous-jacents sont multiples, certains étant encore discutés comme étant ou non d'origine épigénétique (les prions, par exemple). Dans cette partie, seules les marques épigénétiques relatives à la structure chromatinienne seront traitées (**Figure 10**).

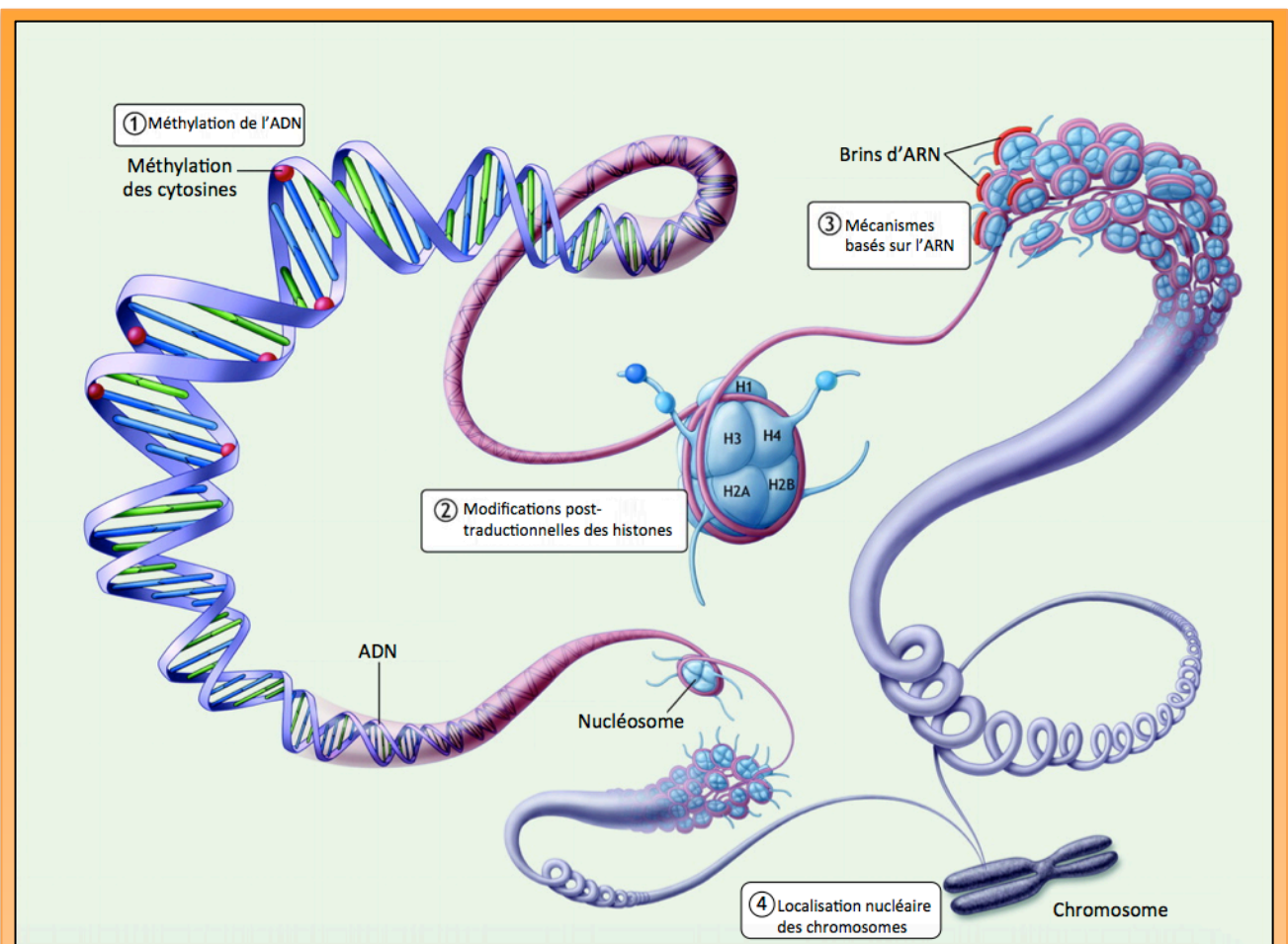
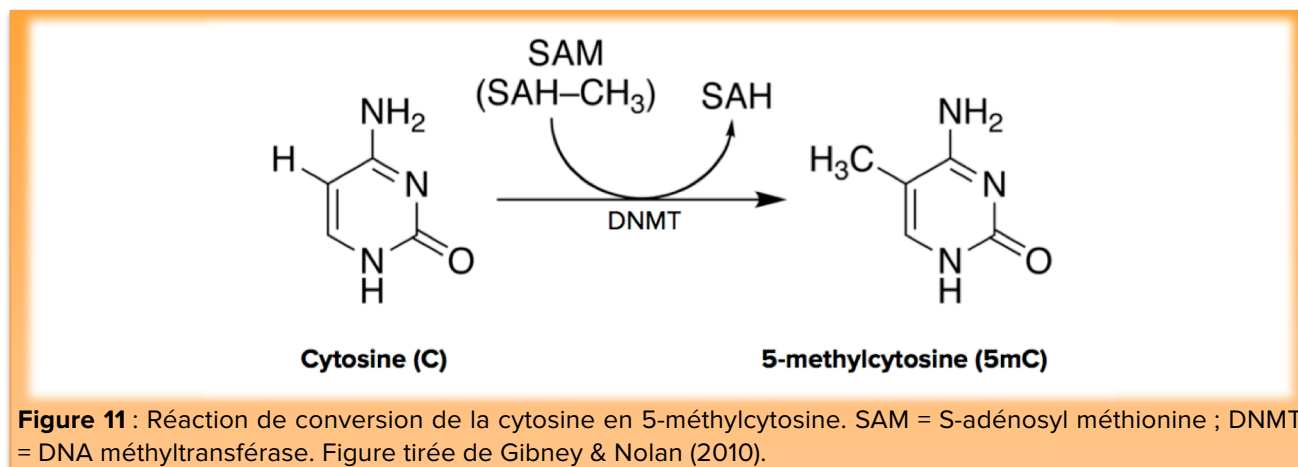


Figure 10 : Représentation graphique de quatre porteurs d'information épigénétique influençant la structure chromatinienne. Figure adaptée de Matouk & Marsden (2008).

1.2.2.1. Méthylation des cytosines

La méthylation de l'ADN correspond généralement à une modification covalente de l'ADN par l'ajout d'un groupement méthyl (CH_3) sur le carbone 5 d'une cytosine, pour former une base appelée 5-méthylcytosine (5mC) (**Figure 11**).



La mise en place de la méthylation se fait par des enzymes de type ADN méthyltransférase (DNMTase). Il existe trois DNMTase principales chez les mammifères (Smith & Meissner 2013). La méthylation dite « de maintenance » permet la conservation des cytosines méthylées lors de la réplication de l'ADN. Elle est catalysée par DNMT1 (Smith & Meissner 2013). On retrouve deux types de DNMT3 impliquées dans la méthylation *de novo*. DNMT3A se charge principalement des régions promotrices et géniques alors que DNMT3B est active dans les séquences répétées et péri-centromériques (Smith & Meissner 2013). DNMT3A et DNMT3B sont assistées par DNMT3L, qui sert de régulatrice, mais ne possède pas en elle-même une activité méthyltransférase (Smith & Meissner 2013). TRDMT1 (anciennement DNMT2) est une autre DNMT, considérée pendant longtemps comme une autre méthyltransférase par homologie de séquence. Toutefois, il a été mis en évidence en 2006 que sa fonction de méthylation ciblait un ARN de transfert (Goll *et al.* 2006) et son rôle en tant que DNMT est aujourd'hui controversé. La présence de ces enzymes de méthylation varie chez les invertébrés. Des orthologues de DNMT1 et DNMT3 sont présents chez l'abeille *Apis mellifera* (Lyko *et al.* 2010; Ruden 2015), le puceron *Acyrtosiphon pisum* (Flores & Amdam 2011; Flores *et al.* 2013) ou l'huître *Crassostrea gigas* (Gavery & Roberts 2010), alors que chez le bombyx du mûrier *Bombyx mori* et le mollusque *Biomphalaria glabrata*, seule DNMT1 est retrouvée (Xiang *et al.* 2010; Fneich *et al.* 2013). Ces enzymes sont absentes chez le nématode modèle *Caenorhabditis elegans* (Gutierrez 2004).

Il existe trois contextes de séquences nucléotidiques dans lesquels la méthylation peut prendre place. Deux de ces contextes sont symétriques sur les deux brins d'ADN : CpG (ou CG selon les auteurs) et CHG (le H pouvant être adénine, thymine ou cytosine). La méthylation non symétrique, présente sur un seul brin, existe sous le contexte CHH. Seule la méthylation symétrique peut être maintenue chez les mammifères lors de la réplication de l'ADN grâce à DNMT1 (Smith & Meissner 2013). Il existe chez les plantes d'autres mécanismes permettant le maintien de la méthylation en contexte CHH lors de la division cellulaire (Chan *et al.* 2005; Du *et al.* 2015).

La proportion et la localisation génomique de la 5-méthylcytosine varie grandement en fonction des phyla (Bird 2002). Chez les métazoaires, la distribution de cette marque fluctue énormément d'une espèce à une autre (Flores & Amdam 2011). Elle est présente chez tous les mammifères et presque exclusivement dans le contexte de dinucléotides CpG. Ces derniers sont répartis sur l'ensemble du génome, et 70% à 80% d'entre eux sont méthylés (Law & Jacobsen 2010). On retrouve régulièrement des regroupements de CpG (îlots CpG) au niveau des régions régulatrices en 5' de gènes (promoteurs). La méthylation de ces îlots dans les promoteurs inhibe la transcription des gènes en aval, en bloquant l'accès à des facteurs de transcriptions ou en facilitant l'attachement de répresseurs (Deaton & Bird 2011). La méthylation en contexte CHH est pratiquement absente des mammifères, à l'exception d'une petite proportion retrouvée dans les cellules souches embryonnaires, et il est présumé qu'elle puisse jouer un rôle dans la différenciation cellulaire (Law & Jacobsen 2010). Chez l'être humain, on estime à environ 1% la proportion de 5-méthylcytosine sur l'ensemble du génome (Bird 2002).

On retrouve un profil de méthylation de type mosaïque chez la plupart des invertébrés car elle n'est pas répartie de façon uniforme dans tout le génome. Elle est essentiellement présente dans les corps de gènes, principalement dans le contexte CpG, et marque souvent les jonctions introns - exons (Schübeler 2015). Certains invertébrés notables pour leur rôle d'organismes modèles en biologie, tels que *Drosophila melanogaster* et *Caenorhabditis elegans*, ont une absence ou quasi absence de méthylation (Bird 2002), alors que chez *Apis mellifera* et *Acyrtosiphon pisum*, elle régule les différents phénotypes développementaux observés chez ces espèces (Hunt *et al.* 2010; Lyko *et al.* 2010).

Les plantes modèles ont pour la plupart un génome très méthylé. Près de 6% des cytosines d'*Arabidopsis thaliana* portent cette marque, indistinctement du contexte CpG, CHG ou CHH. La méthylation survient surtout dans les régions répétées du génome et dans les éléments transposables. Chez les fungi, sa présence est très variable. Elle est absente chez *Saccharomyces cerevisiae*, mais présente chez *Neurospora crassa* avec jusqu'à 2% des cytosines totales étant méthylées (Montanini et al. 2014).

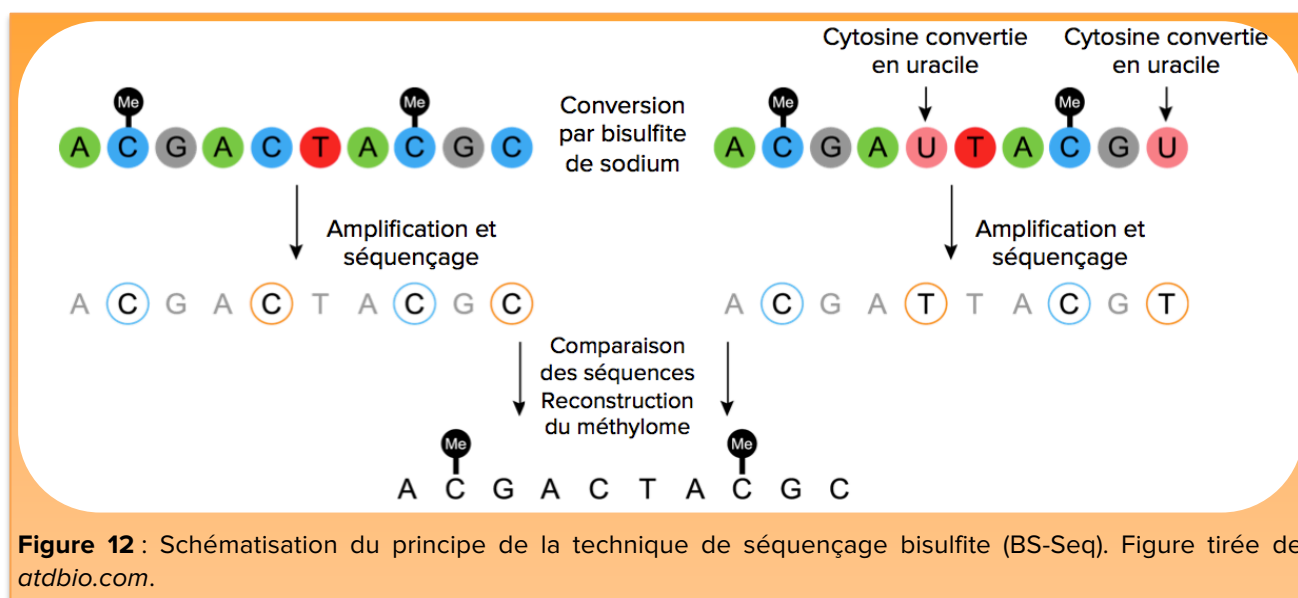
Les fonctions que remplissent la méthylation sont très diverses selon les espèces. Chez les mammifères, elle contrôle principalement la répression de la transcription des gènes lorsqu'elle est présente au niveau des régions promotrices dans les îlots CpG. Elle est impliquée dans la défense des génomes par inactivation (*silencing*) des éléments répétitifs du génome. On la retrouve ainsi très présente dans les éléments génétiques mobiles ainsi que dans les régions centromériques et péricentromériques (He et al. 2011). Chez les mammifères, elle joue un rôle essentiel dans le développement embryonnaire, la différenciation cellulaire, l'inactivation aléatoire du chromosome X, l'empreinte parentale (régions du génome activées ou inactivées par méthylation en fonction de l'origine paternelle ou maternelle de l'allèle transmise) et le vieillissement. Elle est très étudiée en santé humaine car son implication dans le développement de nombreuses maladies à caractère environnemental (ex : associées au tabagisme) est de plus en plus montré (Robertson 2005). En cancérologie, l'hyperméthylation de gènes répresseurs de tumeurs et l'hypométhylation d'oncogènes sont à l'origine du développement de cancer. Chez les invertébrés (tout comme les vertébrés d'ailleurs), on ignore l'action de la méthylation intragénique, mais il a été montré qu'elle module l'épissage alternatif chez les abeilles (Flores & Amdam 2011). L'inactivation des éléments transposables et autres séquences répétitives constitue la fonction principale chez les végétaux (Law & Jacobsen 2010; He et al. 2011). Le rôle de la méthylation chez les fungi est variable. Elle est impliquée dans la transition entre les différents stades développementaux et la défense contre les éléments transposables chez certaines espèces (Jeon et al. 2015). Chez d'autres, l'inactivation par mutation des méthyltransférases n'induit pas de phénotypes visibles (Montanini et al. 2014).

Il existe une héritabilité, somatique et méiotique, de la 5mC. Chez les plantes, la majeure partie de ces marques est transmise à la descendance (Tsaftaris & Polidoros 1999; Peng & Zhang 2009). Des lignées isogéniques stables dont les profils de méthylations sont

transmissibles à la descendance sont d'ailleurs disponibles (Johannes & Colome-Tatche 2011). La gamétogénèse des vertébrés implique un effacement de la 5mC, suivi d'une reméthylation spécifique, essentielle au développement embryonnaire. L'empreinte parentale, par laquelle certaines régions et gènes sont méthylés dans les gamètes en fonction de leur origine paternelle ou maternelle, en est un exemple. Il arrive que certains sites échappent à la déméthylation et soient transmis à la descendance (Bird 2002; Szyf 2015; Prokopuk *et al.* 2015). Le maintien de la méthylation des séquences répétées est un autre exemple d'héritabilité de la 5mC.

Cytosines et 5-méthylcytosines ont la particularité de pouvoir subir une déamination spontanée, les convertissant respectivement en uracile et thymine. Si la présence de l'uracile dans l'ADN génomique est anormale et rapidement réparée enzymatiquement par la cellule (par excision de l'uracile et son remplacement par une cytosine), la thymine persiste sous forme de mutation de transition (Chahwan *et al.* 2010). Il est estimé que la déamination de 5-méthylcytosine survient de 20 à 40 fois plus fréquemment que n'importe quel autre type de mutation ponctuelle sur des nucléotides non-méthylés, et est donc susceptible de jouer un rôle important dans l'évolution des génomes (Russo *et al.* 2012). Ce phénomène est aussi responsable de la dégénérescence des éléments transposables. Ceux-ci sont habituellement hyperméthylés chez la plupart des organismes, et la déamination progressive de leurs 5-méthylcytosines finit par modifier leur séquence jusqu'à ce qu'ils deviennent complètement inactivés (Jones 2012; Wang *et al.* 2013).

La principale technique permettant d'obtenir une vue d'ensemble des cytosines méthylées sur la totalité d'un génome est celle du séquençage bisulfite (BS-Seq). Pour utiliser cette méthode, l'ADN extrait d'un organisme, tissu, ou lignée cellulaire est soumis à un traitement au bisulfite de sodium. Ce traitement convertit les cytosines non-méthylées en uracile (**Figure 12**). L'ADN converti est alors amplifié par PCR, ce qui a pour effet de remplacer les uraciles par des thymines, et est ensuite séquencé. Les 5mC ne sont pas converties par le bisulfite, et sont ainsi les seules observées dans les séquences d'ADN traitées. Avec l'aide d'outils bioinformatiques, on peut reconstruire le méthylome, avec une résolution au nucléotide près, et obtenir le pourcentage de méthylation à un site donné dans une population cellulaire (la méthylation pouvant fluctuer selon les cellules, tissus, étapes du développement, etc.).



1.2.2.2. Autres modifications chimiques des nucléotides

Plusieurs autres variants de la cytosine sont retrouvés dans l'ADN génomique (**Figure 13**). La 5-formylcytosine (5fC) et la 5-carboxylcytosines (5caC) sont des intermédiaires secondaires du processus de déméthylation des 5mC. Pour l'instant, on ignore si ces deux nucléotides disposent d'une position génomique préférentielle, mais leur accumulation dans certains tissus ou lignées cellulaires (cerveau et cellules souches embryonnaires chez les souris par exemple) laisse présager qu'ils puissent avoir une fonction régulatrice. En l'absence actuelle de protéines connues pour lire ces marques ou s'y fixer spécifiquement, cela reste encore à démontrer (Song *et al.* 2012). Autre intermédiaire de la déméthylation, la 5-hydroxyméthylcytosine (5-hmC) se localise avec une forte affinité aux sites d'initiation de la transcription (TSS), promoteurs, exons et amplificateurs (*enhancers*), et on présume une action dans la régulation de l'expression génique (Song *et al.* 2012). Tout comme la 5fC et la 5caC, la 5hmC se retrouve en plus grande proportion dans les cellules souches embryonnaires et les tissus cérébraux des mammifères (Song *et al.* 2012). Elle interviendrait aux stades précoces du développement embryonnaire. Sa présence a été confirmée dans le génome de la plupart des invertébrés séquencés (Pastor *et al.* 2013). Les enzymes *Ten-Eleven-Translocations* (TET) catalysent les réactions entre la 5mC et ces trois bases intermédiaires (Song *et al.* 2012).

La cytosine n'est pas l'unique nucléotide pouvant être modifiée dans l'ADN. La N⁶-méthyladénine (m⁶A) est communément retrouvée chez les procaryotes. Avec la 5mC, ces deux nucléotides méthylés servent à protéger l'ADN bactérien de la digestion par enzymes

de restrictions présentes dans le cytoplasme; ces dernières ciblant des ADNs exogènes, tels que ceux des phages (Ratel *et al.* 2006). La m⁶A module également l'expression génique et la virulence (Low *et al.* 2001). Chez *Chlamydomonas reinhardtii*, 0,4% des adénines sont méthylées, principalement au niveau des promoteurs géniques et des sites d'initiation de la transcription. Sa présence est corrélée avec une activité transcriptionnelle plus importante (Fu *et al.* 2015). Elle n'existe qu'à l'état de trace chez les eucaryotes supérieurs, et sa fonction est encore mal connue (Ratel *et al.* 2006). Cette base est faiblement présente (0,025%) chez *Caenorhabditis elegans*, mais contribue à des effets transgénérationnels (Greer *et al.* 2015). Un pourcentage légèrement supérieur (0,07%) de m⁶A est retrouvé chez *Drosophila melanogaster*, essentiellement au niveau des éléments transposables et de certains sites spécifiques lors des étapes précoces de développement, mais sa fonction exacte n'est pas encore élucidée (Zhang *et al.* 2015).

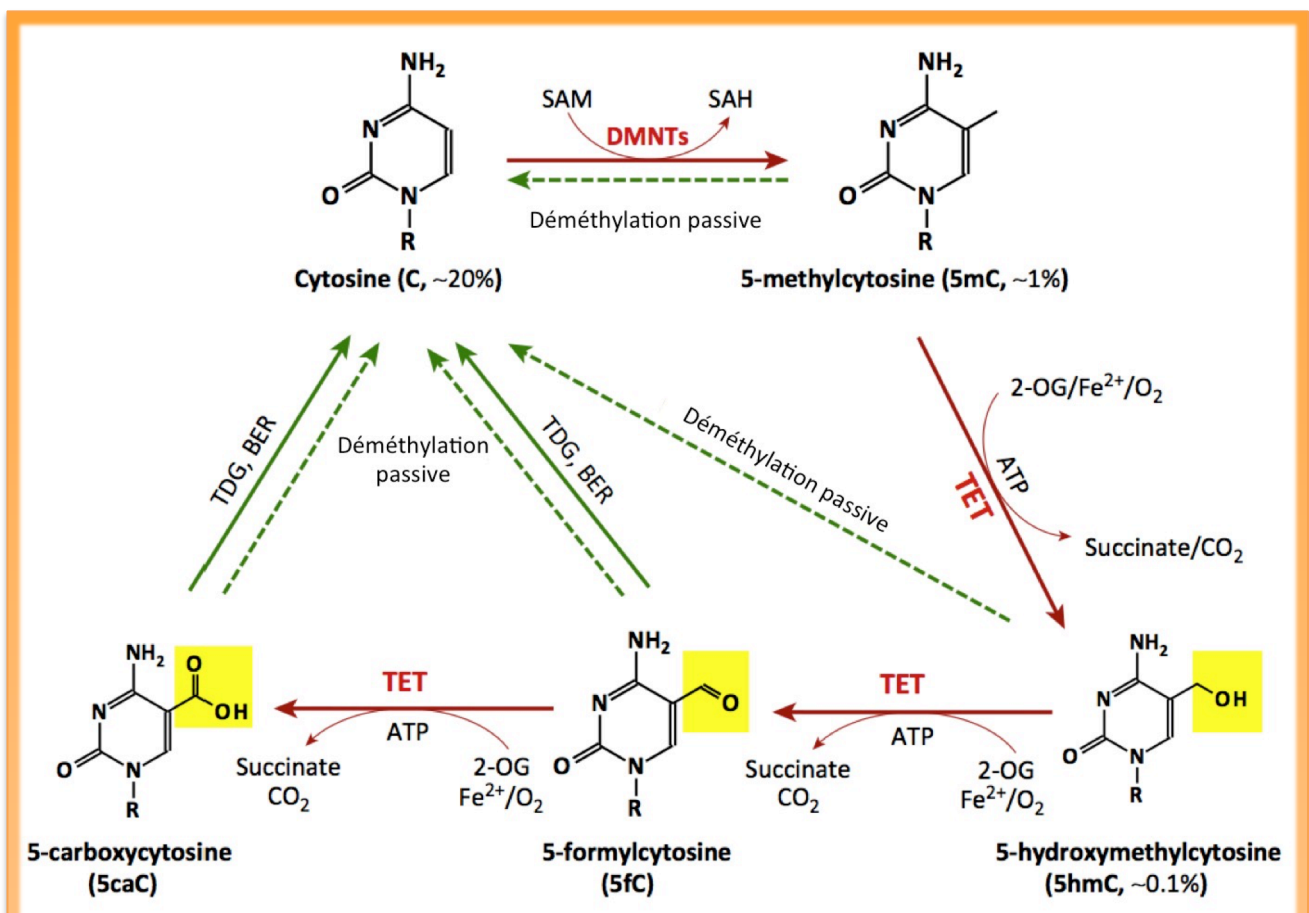
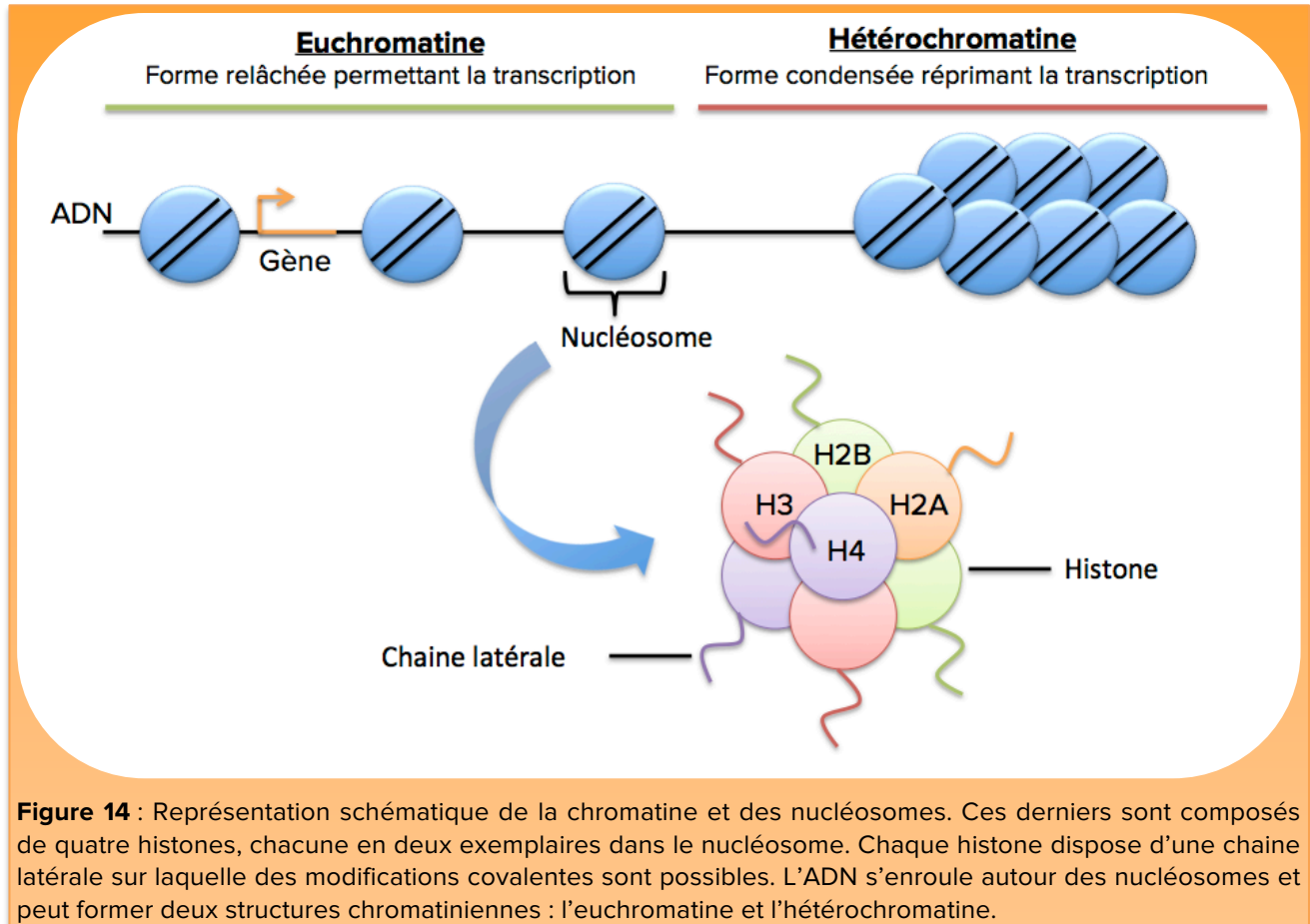


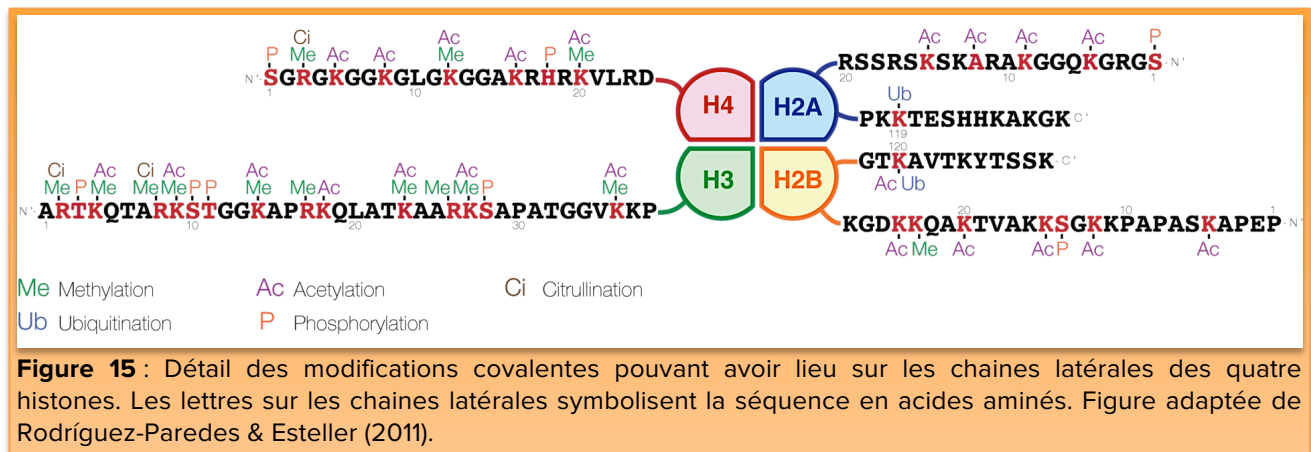
Figure 13 : Cytosine (C), 5-méthylcytosine (5mC), et les trois intermédiaires de la déméthylation de la 5mC : la 5-hydroxyméthylcytosine (5hmC), 5-formylcytosine (5fC) et 5 carboxylcytosine (5caC), ainsi que leur proportion dans le génome des mammifères. TET = Ten-Eleven Translocation ; TDG = Thymine DNA glycosylase ; ATP = Adénosine Tri Phosphate ; BER = Base Excision Repair ; SAM = S-Adénosyl Méthionine ; SAH = S-Adénosyl Homocystéine ; DNMTs = DNA méthyltransférases. Figure tirée de Huang & Rao (2014).

1.2.2.3. Modification des histones et position des nucléosomes

Dans le noyau cellulaire, l'ADN n'est pas isolé et interagit avec de nombreuses protéines nucléaires pour former la chromatine. Plus spécifiquement, l'ADN se complexe avec des octamères d'histones (deux histones H2A, deux H2B, deux H3 et deux H4), en s'enroulant autour d'eux sur une longueur de 147 paires de bases, pour former un nucléosome (**Figure 14**).



Les quatre types d'histones possèdent des chaînes latérales (ou queues) accessibles à l'extérieur du nucléosome et dont certains acides aminés sont candidats à des modifications post-traductionnelles covalentes (**Figure 15**). Une méthylation peut avoir lieu sur les lysines et arginines, acétylation et ubiquitylation sur les lysines uniquement, phosphorylation sur les sérines, thréonines et tyrosines, et citrullination sur les arginines. En tout, près d'une centaine de modifications chimiques est possible (Rodríguez-Paredes & Esteller 2011). Les histones et leurs modifications sont très conservées chez les eucaryotes, indiquant un rôle important. Les dinoflagellés (algues unicellulaires) sont les seuls eucaryotes connus dépourvus d'histones (Rizzo 2003).



Les modifications biochimiques des histones sont catalysées par différentes enzymes relativement conservées chez les eucaryotes. Impliquées dans le marquage des histones, on trouve les arginines déiminases (PADs, *Protein Arginine Deiminase*), arginines méthyltransferases (PRMTs, *PRotein arginine MethylTranferases*), lysines méthyltransferase, lysines acetyltransferases (KATs), lysines biotinases (HLCS), lysines ribosylases, lysines ubiquitinases, et certaines sérines/thréonines/tyrosines kinases (Zhang & Reinberg 2001). Les acétyltransférases (HATs) sont parmi les plus étudiées (Peserico & Simone 2011). Elles se répartissent en quatre familles impliquées dans l'activation de la transcription : GCN5, MYST, P300/CBP et coactivateurs des récepteurs nucléaires SRC/P160. On retrouve plusieurs types d'enzymes dont la fonction est de retirer des modifications chimiques des histones : arginines déméthylases, lysines déacétylases (HDACs), lysines déméthylases, lysines déribosylases, lysines déubiquitinases et sérines/thréonines/tyrosines phosphatases. Les HDACs ont des fonctions antagonistes aux HAT et se divisent en cinq classes : classe I comprenant HDAC1, 2, 3 et 8, classe IIa comprenant HDAC 4, 5, 7 et 9, classe IIb comprenant HDAC 6 et 10, classe III comprenant les sirtuins et classe IV comprenant les HDAC11 (Peserico & Simone 2011). L'inactivation des HDACs peut avoir différents effets sur l'organisme. Chez l'homme, certains inhibiteurs de HDACs sont utilisés de manière thérapeutique pour le traitement de troubles émotifs, de l'épilepsie ou de neurodégénérescence (Hahnen *et al.* 2008), mais peuvent également causer la réactivation de retrovirus latents (Arbuckle & Medveczky 2011).

Les fonctions jouées par les modifications d'histones vont dépendre du (I) type de modification chimique, (II) de la présence d'autres types de modifications au même endroit, ainsi que (III) de la localisation génomique de la marque.

I) Par exemple, la triméthylation de la lysine 27 de l'histone 3 (H3K27me3) est une marque répartie sur de grandes régions du génome (Trojer & Reinberg 2007; Bannister & Kouzarides 2011), et associée à la répression de la transcription. L'acétylation de la même lysine (H3K27ac) est positionnée de manière très ponctuelle sur les amplificateurs qu'elle active (Tie *et al.* 2009; Creyghton *et al.* 2010; Liu *et al.* 2011; Shlyueva *et al.* 2014).

II) Lorsqu'elle est présente sur le corps de gènes, la triméthylation de la lysine 36 de l'histone 3 (H3K36me3) est un signe de gènes activement transcrits (Barth & Imhof 2010). En dehors des gènes, elle sert à marquer des sites de mésappariement de nucléotides pour activer un mécanisme de réparation (Li *et al.* 2013).

III) Au site d'initiation de la transcription (TSS), H3K4me3 (triméthylation de la lysine 4 de l'histone 3) est indicative de gènes transcriptionnellement compétents (Berger 2007; Gu & Fire 2009; Kharchenko *et al.* 2010; Zentner & Henikoff 2013). Cependant, chez les cellules souches embryonnaires des vertébrés, une colocalisation au même endroit avec H3K27me3 indique plutôt des gènes impliqués dans la différenciation cellulaire et dont la transcription est suspendue (Bernstein *et al.* 2006; Pan *et al.* 2007; Vastenhouw & Schier 2012).

Globalement ce n'est pas une marque, mais une combinaison de marques dans un contexte génomique donné (Filion *et al.* 2010; Barth & Imhof 2010) qui influera sur la structure chromatinienne. Cette dernière se trouvera ainsi dans des états plus ou moins condensés dont les deux formes extrêmes seront l'euchromatine et l'hétérochromatine. L'euchromatine est un état décondensé de l'ADN et permissif à la transcription, alors que l'hétérochromatine est une forme compactée dans laquelle la transcription est réprimée (**Figure 14**). Les différents états de compaction de la chromatine sont réversibles et peuvent passer de l'un à l'autre en fonction du niveau de méthylation de l'ADN et des formes d'histones présentes (Berger 2007).

On retrouve aussi des variants des histones H3 et H2, tout aussi conservés chez les eucaryotes que leurs versions canoniques (Talbert & Henikoff 2010; Nashun *et al.* 2010), mais avec des fonctions différentes. Par exemple, CenH3 est spécifiquement localisée dans les centromères des chromosomes et H2A.X indique des sites de cassure du double brin d'ADN (Talbert & Henikoff 2010).

Outre les modifications chimiques des histones, la position des nucléosomes le long d'une séquence d'ADN a aussi son importance. Les gènes compétents à la transcription se caractérisent par une absence de nucléosomes autour du TSS, ce qui réduit l'encombrement stérique et favorise l'assemblage de la machinerie transcriptionnelle (Jiang & Pugh 2009).

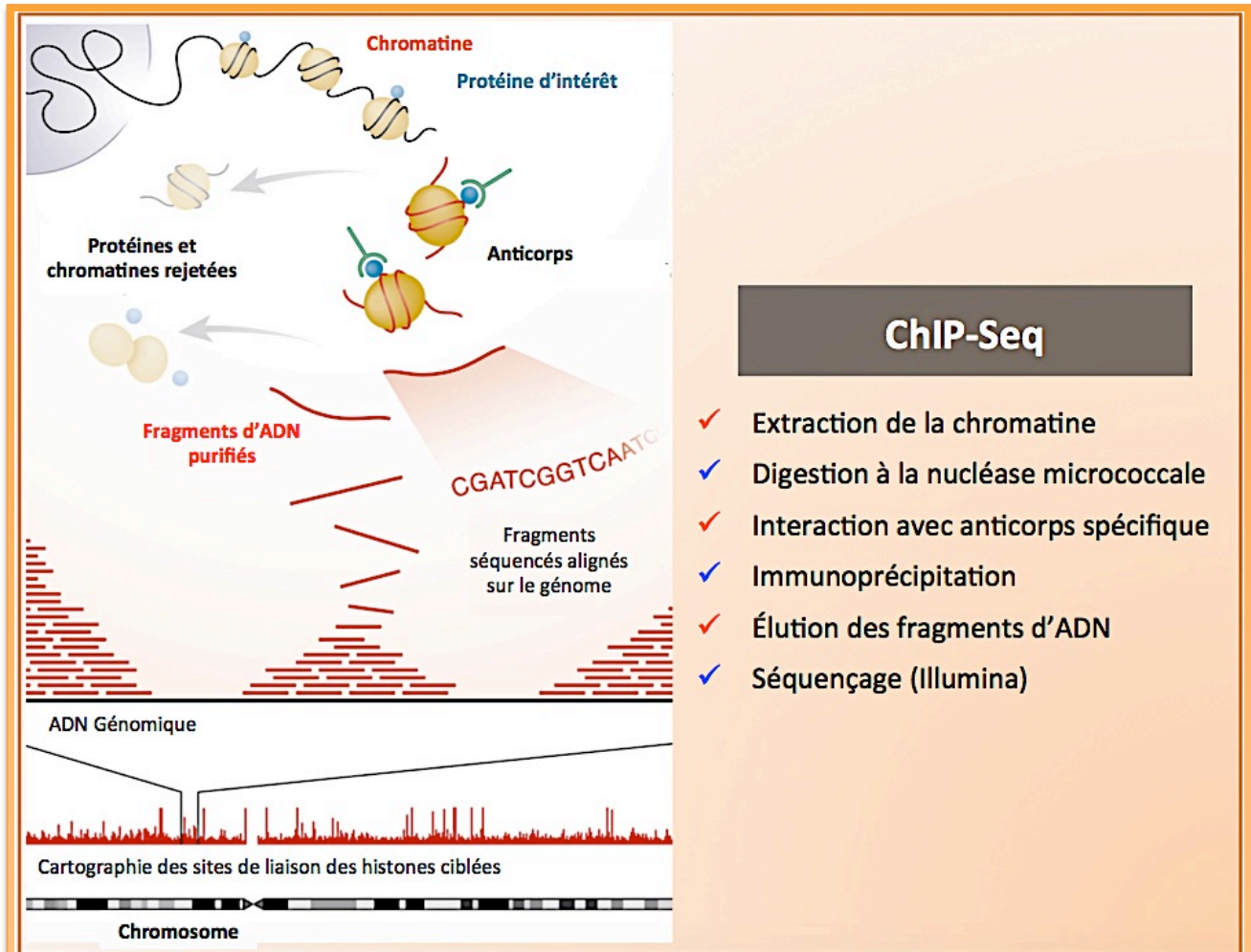


Figure 16 : Représentation schématique de la technique du ChIP-Seq et de ses étapes clefs. Figure tirée de Botcheva *et al.* (2011).

La technique d'immunoprécipitation de chromatine suivi de séquençage (ChIP-Seq) est la principale employée pour l'étude de la répartition génomique des modifications d'histones (**Figure 16**). Cette méthode étant utilisée à plusieurs reprises au cours de ma thèse, il me semble nécessaire de la décrire en détail. Elle permet d'identifier des régions génomiques sur lesquelles sont fixées des protéines cibles. Dans un contexte d'études sur l'épigénétique, ce sont généralement les séquences d'ADN enroulées autour des nucléosomes portant les modifications d'histones qui sont visées. Une étape de *crosslink*

peut être effectuée pour accroître les interactions en liant de manière covalente l'ADN et les protéines, mais n'est normalement pas nécessaire lorsque l'on travaille sur les histones car ces dernières sont étroitement liées à l'ADN. Dans un ChIP natif (sans *crosslink*), la chromatine extraite est digérée par la nucléase micrococcale (MNase), une enzyme bactérienne coupant l'ADN spécifiquement entre les nucléosomes. On utilise ensuite un anticorps, ciblant une histone ou une modification chimique d'histone. Le complexe anticorps/histone/ADN est immunoprécipité par ajout de microbilles de sépharose recouvertes de protéine A. Cette dernière a une très forte affinité avec les anticorps, et permet par conséquent de piéger le complexe. On procède ensuite à l'élution des fragments d'ADN immunoprécipités, qui sont alors séquencés de manière massive (ex : séquençage Illumina) et repositionnés sur le génome de référence. Leur distribution sur le génome indique l'emplacement des histones / modification d'histones ciblées.

I.2.2.4. ARNs non-codants

Avec l'explosion des technologies de séquençage massif au début des années 2000, la transcriptomique a pris une ampleur sans précédent. Grâce à la sensibilité de ces instruments, il a été possible de mettre en évidence des ARNs non-codants (non traduits en protéines), mais impliqués dans la régulation de la structure chromatinienne des eucaryotes (Holoch & Moazed 2015). Le mécanisme d'interférence ARN (RNAi), retrouvé chez presque tous les eucaryotes, se caractérise par de petits ARNs non-codants. Il permet la répression de l'expression génique, et module la méthylation de l'ADN et les modifications d'histones (Holoch & Moazed 2015). Plus d'une cinquantaine d'autres types d'ARNs non-codants sont documentés. Citons entre autres les *micro RNAs* (miRNA), les *small interfering RNAs* (siRNA), les *small nucleolar RNAs* (snoRNA), les *X inactive specific transcript RNAs* (XistRNA) et les *long non coding RNAs* (lncRNA) (Cech & Steitz 2014).

Chaque type d'ARN occupe généralement une fonction très spécifique, mais leur activité principale tourne autour de l'extinction post-transcriptionnelle de gènes (*post transcriptionnal gene silencing*, PTGS). Cette action peut se répercuter par des modifications de la structure chromatinienne facilitant la répression de l'expression sur le long terme (*i.e.* hétérochromatisation). Chez les plantes, mais aussi les animaux, ce mécanisme est particulièrement utilisé pour inactiver les éléments transposables (Slotkin &

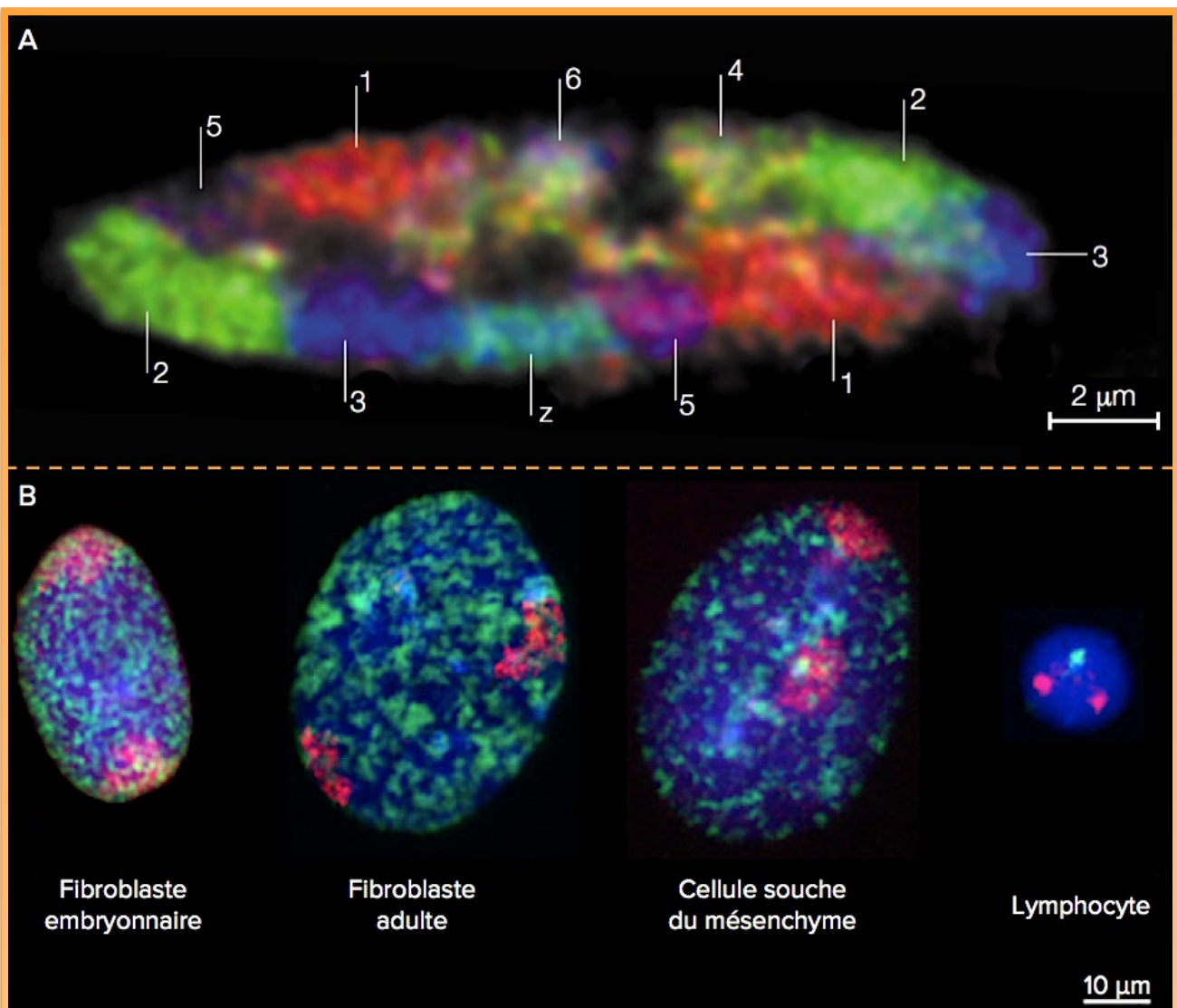
Martienssen 2007). Dans les cellules embryonnaires des femelles mammifères, les XistRNA recouvrent en totalité un des deux chromosomes X pour induire son inactivation et sa méthylation. Les lncRNAs, quant à eux, agissent plutôt au niveau de gènes ciblés qu'à l'échelle du chromosome, et peuvent provoquer l'activation ou la répression de leur cible (Holoch & Moazed 2015). Enfin, les siRNAs sont des acteurs majeurs des boucles autocatlytiques. Elles assurent le maintien de l'information épigénétique sur certains loci par renforcement de la méthylation de l'ADN et des modifications d'histones typiques de l'hétérochromatine (Holoch & Moazed 2015).

1.2.2.5. Localisation nucléaire des chromosomes

Au sein du noyau des eucaryotes, chromosomes et séquences d'ADN ne sont pas disposés aléatoirement, et se déplacent en fonction du cycle cellulaire et de l'activité de la cellule. Par fusion transgénique de protéines fluorescentes avec des histones, ou par hybridation fluorescente *in situ* (FISH) avec des sondes ADN ciblant des séquences présentes spécifiquement sur un ou plusieurs chromosomes, on peut remarquer que chaque chromosome occupe un espace défini de forme elliptique et dont l'arrangement se structure radialement au noyau (**Figure 17 A**) (Cremer & Cremer 2001; Meaburn & Misteli 2007; Cremer & Cremer 2010). Ce positionnement fluctue selon les types cellulaires et tissulaires (Foster *et al.* 2012; Voldgorn *et al.* 2015) (**Figure 17 B**), mais se trouve conservé évolutivement (Meaburn & Misteli 2007), raison pour laquelle on considère ce phénomène comme étant de nature épigénétique. Les chromosomes riches en gènes vont avoir une position plus centrale et généralement associée à des structures euchromatiques (permettant la transcription) plus importantes que ceux pauvres en gènes, situés en périphérie et avec une plus forte proportion d'hétérochromatine réfractaire à la transcription (Tanabe *et al.* 2002). Cette organisation est dynamique et réversible, et les séquences géniques peuvent se déplacer de la périphérie vers l'intérieur du noyau, ou sortir de leur compartiment chromosomique, ce qui s'accompagne souvent de la formation de structure euchromatique propice à l'expression génique (Meaburn & Misteli 2007). Le même phénomène est observable au cœur du territoire d'un chromosome, au sein duquel les séquences plus centrales seront principalement inactives, et celles en périphéries plus susceptibles d'être transcrites ou d'interagir avec les séquences d'autres chromosomes adjacents (ex : translocation) (Boyle *et al.* 2011). La seule exception connue est la levure

Saccharomyces cerevisia au stade bourgeonnant, pour laquelle aucune localisation chromosomique spécifique n'a pu être mise en évidence (Cremer & Cremer 2010).

Plusieurs facteurs influencent la position des chromosomes dans la cellule : différenciation cellulaire (Foster et al. 2012, Voldgorn et al. 2015), processus cellulaires et moléculaires fondamentaux, tels que des modifications de la transcription (Cremer & Cremer 2001; Berezney 2002), ou des dégradations de l'ADN (cassures, mésappariements de bases) et leurs réparations (Mehta et al. 2013). La position nucléaire des chromosomes est aussi liée au cancer. Des territoires des chromosomes anormaux peuvent être la cause de translocations menant à cette pathologie. De la même manière, la plupart des cellules cancéreuses ont des territoires des chromosomes distincts (Bourne et al. 2013).



I.2.3. Liens entre *Dauermodifikationen* et épigénétiques

L'épigénétique peut donc être décrite comme un ensemble de modifications réversibles et héritable, mais non codées dans la séquence nucléotidique d'ADN, et qui est capable d'influencer le phénotype (généralement en influençant la transcription). Il peut s'agir par exemple de modifications chimiques de l'ADN (ex : méthylation des cytosines), de protéines interagissant avec l'ADN (ex : modification des histones), de petits ARNs non-codants, ou de la localisation nucléaire des chromosomes (Wu & Morris 2001). Divers facteurs environnementaux, tels que la température, la pollution, la présence d'autres individus, de prédateurs ou la nutrition peuvent avoir une influence sur les mécanismes de régulation épigénétique (Feil & Fraga 2012). Alors que la séquence nucléotidique d'ADN code pour la composition et la fonction des gènes, l'information épigénétique permet un contrôle du moment et du contexte dans lequel ces gènes peuvent être exprimés. Le phénotype se définit alors comme le produit de la triple interaction entre génome, épigénome, et environnement. Ainsi, les modifications de la séquence d'ADN ne sont pas les seuls changements pouvant générer de nouveaux phénotypes. Deux génomes identiques, sous des conditions environnementales différentes et avec des épigénomes distincts peuvent présenter des phénotypes variés.

Les *Dauermodifikationen* ne peuvent pas être expliquées simplement par des changements génétiques, à cause de la rapidité avec laquelle elles apparaissent (parfois au sein d'une même génération, comme chez *Daphnia*) et leur métastabilité (*i.e.* réversibilité après un certain laps de temps). Si à l'époque de Victor Jollos, l'idée était complètement réfutée, il est maintenant clair qu'il existe des sources non génétiques de phénotypes héritable, et en particulier les modifications épigénétiques (Nilsson *et al.* 2012; Skinner *et al.* 2014; Nilsson & Skinner 2015; Ruden 2015; Prokopuk *et al.* 2015). En réponse à des changements d'environnement, on sait qu'elles sont en mesure d'induire de nouveaux phénotypes, et de permettre la sélection de ce phénotype par l'environnement en gardant le génotype constant. Les modifications épigénétiques sont donc d'excellents candidats pour expliquer les *Dauermodifikationen*. Il faut toutefois bien caractériser comment et à quel moment elles agissent en réponse à l'environnement, et dans quelles mesures elles peuvent être héritable, somatiquement (lors du développement par exemple) et/ou méiotiquement.

1.2.3.1. Caractéristiques d'un modèle d'étude *Dauermodifikationen*/épigénétique

Pour établir des liens clairs entre *Dauermodifikationen* et modifications épigénétiques, il est nécessaire de travailler sur des modèles biologiques avec des caractéristiques appropriées. J'ai défini trois critères qui me semblent essentiels à cet objectif.

I) Le premier est de toute évidence de travailler sur un organisme pour lequel il est facile d'induire des *Dauermodifikationen* en laboratoire. Un modèle avec temps de génération court offrirait également l'opportunité d'étudier si les phénotypes induits et les modifications épigénétiques possiblement liés sont transmissibles à la descendance

II) Un autre aspect particulièrement important est la possibilité d'obtenir des individus génétiquement identiques (clonaux). Ce point est essentiel pour être en mesure de se concentrer strictement sur l'épigénétique en éliminant les variations génétiques. La séquence d'ADN en elle-même peut avoir un impact en *cis* sur les modifications épigénétiques (Davidson 2010). De plus, il devient alors beaucoup plus compliqué d'établir un lien direct entre des phénotypes induits par l'environnement et les implications respectives de la génétique et de l'épigénétique. En utilisant des organismes ayant une phase de multiplication asexuée, clonale, on réduit au maximum les variations génétiques inter-individus et donc leur possible impact sur le phénotype. Certains travaux ont montré une part d'hétérogénéité chez des individus théoriquement clonaux (Zhou *et al.* 2000; Bayne & Grevelding 2003; Saravanan & Godhe 2010; Li *et al.* 2015), mais l'impact sera toutefois limité et beaucoup plus facilement mesurable.

III) Les espèces en interactions symbiotiques font de bons modèles pour l'étude de l'évolution adaptative. Les deux partenaires, co-évoluant ensemble pour maintenir leur relation, sont en général réactifs aux changements environnementaux à cause de la forte pression évolutive qui pèse sur eux (Paracer & Amhadjian 2000; Thrall *et al.* 2007; Tellier *et al.* 2014). Les deux extrêmes des relations symbiotiques, mutualismes et parasitismes font donc un bon contexte pour observer l'impact de changement d'environnement, car les deux partenaires doivent être en mesure de réagir rapidement pour maintenir la relation.

Deux modèles étudiés dans mon laboratoire d'accueil remplissent ces trois conditions : le parasite humain *Schistosoma mansoni*, et le corail tropical *Pocillopora damicornis*.

I.3. SCHISTOSOMA MANSONI, UN PARASITE HUMAIN

I.3.1. Description du modèle

I.3.1.1. Cycle de vie

Schistosoma mansoni est un parasite *Plathelminthe* (vers plat) *Trematoda* (appareil digestif avec cæcum) ayant un cycle de vie dixène gonochorique (déterminisme sexuel chromosomique), avec deux hôtes obligatoires et successifs (**Figure 18**). On retrouve un hôte intermédiaire mollusque d'eau douce, du genre *Biomphalaria*, dans lequel le parasite se multiplie de façon asexuée, et un hôte définitif vertébré (rongeur ou primate, incluant l'espèce humaine) à l'intérieur duquel aura lieu une reproduction sexuée. Le parasite traverse ainsi trois environnements notables durant son cycle de vie : organismes des hôtes intermédiaires et définitifs, et milieu aquatique dulçaquicole.

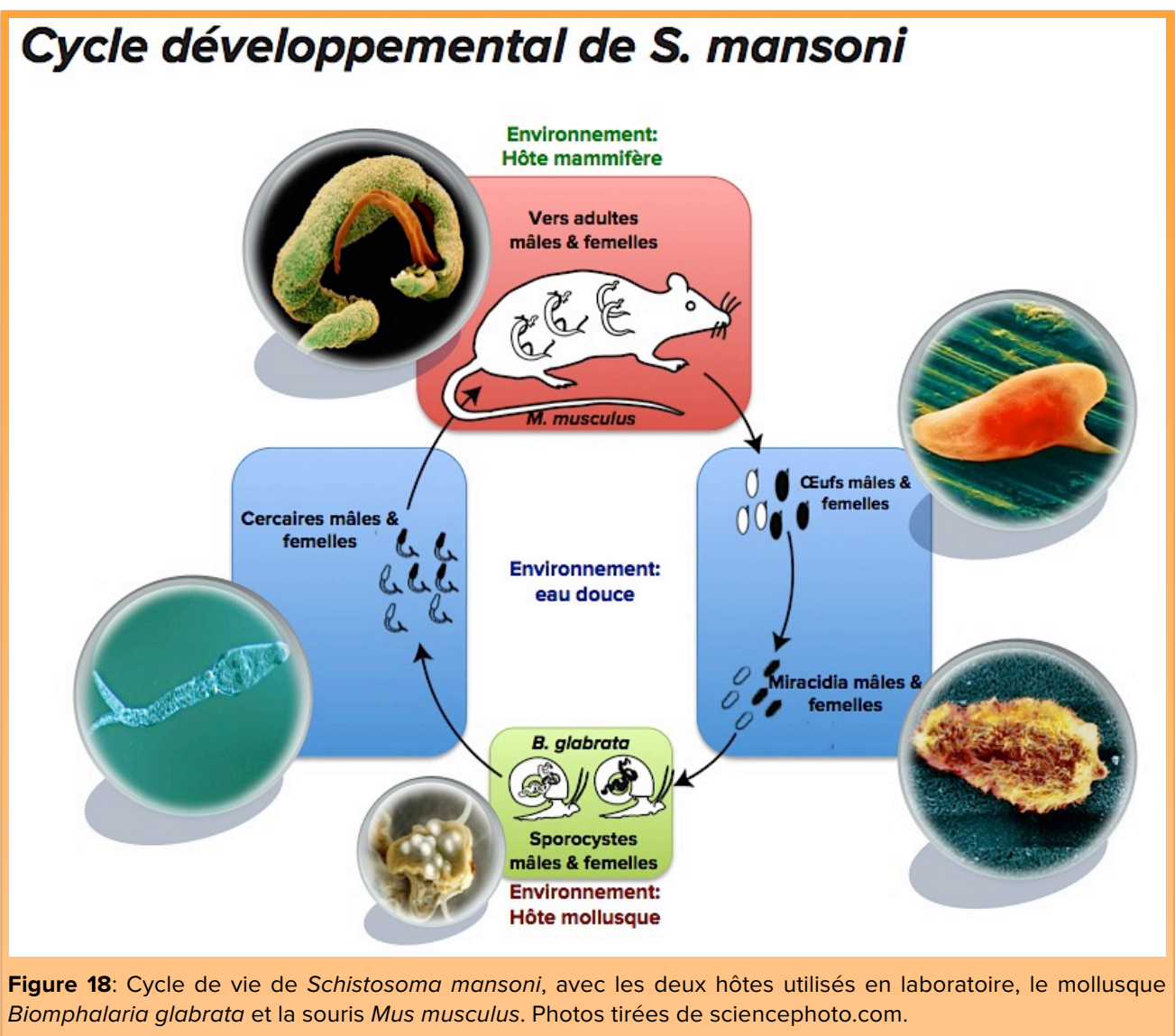


Figure 18: Cycle de vie de *Schistosoma mansoni*, avec les deux hôtes utilisés en laboratoire, le mollusque *Biomphalaria glabrata* et la souris *Mus musculus*. Photos tirées de sciencephoto.com.

Ce cycle commence lorsque des œufs, libérés dans les matières fécales de l'hôte définitif, entrent en contact avec l'eau douce. À l'intérieur des œufs se trouvent une première larve natatoire, le miracidium. Elle est entièrement formée lors de la ponte et ne nécessite pas de développement supplémentaire à l'intérieur de l'œuf. La différence de pression osmotique entre le corps de l'hôte définitif et l'eau douce induit l'éclosion du miracidium. Ce dernier part à la recherche d'hôtes mollusques du genre *Biomphalaria*. Une fois trouvé, le miracidium traverse le tégument de l'hôte, perd ses plaques ciliées, et entame une transformation vers un premier stade intramolluscal, nommé sporocyste primaire (Sp1, ou sporocyste mère). Durant une dizaine de jours, le Sp1 génère des centaines de sporocystes secondaires (Sp2, ou sporocystes filles) qui migrent vers l'hépatopancréas du mollusque. Le développement des Sp2 est associé à une perte complète de fertilité du mollusque. Chaque sporocyste secondaire produit durant son existence plusieurs centaines de cercaires, une seconde larve natatoire. Le processus de multiplication asexuée dans l'hôte mollusque est très efficace, puisqu'un seul miracidium peut être la source de plusieurs milliers de cercaires (Theron 1986). Vingt-huit à trente-cinq jours sont nécessaires entre l'infection par le miracidium et l'émission des premières cercaires. Un mollusque infecté peut émettre des cercaires pendant plusieurs mois avant de mourir.

Une fois émises par les sporocystes, les cercaires migrent dans les tissus du mollusque pour atteindre l'eau. L'espérance de vie d'une cercaire ne dépasse que rarement 24h et son infectivité est limitée à quelques heures durant lesquelles elle cherche activement un hôte vertébré (Théron & Pointier 1985). La queue bifide se scinde de la tête lors de la pénétration du parasite, et seule cette dernière traverse le derme de l'hôte, ce qui initie sa transformation en schistosomule.

La schistosomule, un autre stade larvaire, rejoint la circulation sanguine et atteint le cœur après 24h. Elle entame ensuite une migration vers les poumons qu'elle atteint 5 à 7 jours après la pénétration, puis la veine porte hépatique après 15 jours. C'est lors de ces étapes qu'apparaît le dimorphisme sexuel et c'est à cet endroit que s'effectue l'appariement entre mâles et femelles. Le couple se déplace ensuite vers les veines mésentériques, où il restera jusqu'à sa mort. La durée de vie d'un couple adulte n'est pas connue avec exactitude, mais il est suspecté qu'il puisse survivre plusieurs années (Coura

et al. 1974). La femelle se loge à l'intérieur du canal gynécophore du mâle et n'en sort que pour rejoindre les veines capillaires mésentériques où elle effectue sa ponte, qui peut se chiffrer à 300 œufs quotidiennement. Les œufs sont dotés d'un éperon latéral qui leur permet de franchir la paroi intestinale et d'être expulsés avec les fèces. La position de l'éperon permet d'ailleurs de différencier *S. mansoni* des autres espèces de schistosomes humains. Il faut compter au minimum 32 jours post-infection avant la ponte des premiers œufs. Les adultes mâles font environ 10 mm de longueur, pour 10 à 14 mm pour les femelles (Bogitsh et al. 2013).

1.3.1.2. La bilharziose, maladie tropicale négligée aux lourdes conséquences

Schistosoma mansoni est responsable de la bilharziose, ou schistosomiase intestinale. Cette maladie parasitaire afflige 67 millions de personnes dans le monde (plus de 200 millions pour toutes les formes de schistosomiasés), principalement en Afrique, Amérique du Sud, Caraïbes et Moyen-Orient (Oliveira et al. 2004; King 2010) (**Figure 19**). La maladie est fréquente dans les zones rurales sans assainissement des eaux usées, et les plans d'eau contenant des matières fécales humaines sont à haut risque infectieux.

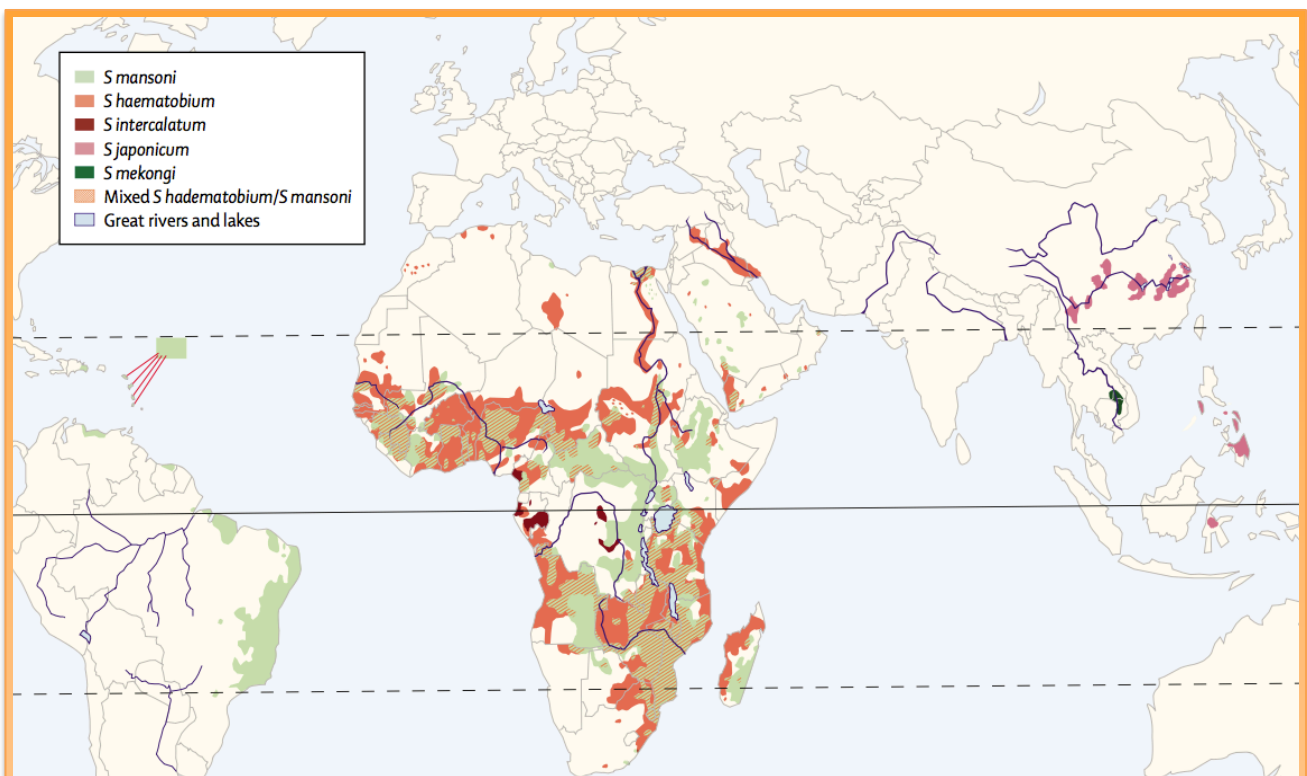


Figure 19 : Répartition géographique des différentes espèces de schistosomes infectieuses pour l'homme. Figure tirée de Gryseels et al. (2006).

La principale cause de la pathologie résulte de l'enkystement des œufs dans différents tissus de l'hôte. Environ la moitié des 300 œufs pondus quotidiennement par une femelle adulte ne traverse pas la paroi intestinale pour rejoindre le milieu extérieur, et est coincée dans les tissus adjacents ou happée par la circulation sanguine et filtrée par le système porte hépatique. Au niveau de l'intestin et du foie, l'accumulation d'œufs induit une réaction hyper-inflammatoire provoquée par un afflux d'éosinophiles, de macrophages et de lymphocytes, ce qui se traduit localement par l'apparition de granulomes. La réversibilité des symptômes inflammatoires provoqués par les œufs dépend de la charge parasitaire et du délai de traitement de l'individu. Si l'infection persiste, la pathologie peut dégénérer en fibrose hépatique et hépatosplénomégalie. Il existe une forme aigüe de la maladie, nommée fièvre de Katayama, susceptible d'apparaître trois à six semaines après l'infection. Cette forme virulente n'est cependant pas la plus courante (Jensen *et al.* 1995; Jamison *et al.* 2006).

Les infections à *S. mansoni* sont rarement mortelles en tant que telles, mais ont un impact socio-économique lourd sur les populations touchées (King 2010). Le praziquantel est actuellement l'unique traitement plébiscité contre la schistosomiase. Cette molécule est efficace sur les vers adultes, et deux doses de 40 à 60 mg/kg de poids de l'hôte, séparées par un mois d'intervalle permettent une guérison dans plus de 90% des cas (Ferrari *et al.* 2003). D'autres schistosomicides, tels que le niridazole, l'hycanthone et l'oxaminiquine ont été utilisés au cours du siècle dernier, mais ont tous été abandonnés par manque d'efficacité, du fait d'une toxicité trop importante pour l'hôte, ou de l'apparition de résistance chez le parasite (da Rocha Pitta *et al.* 2013). De multiples campagnes de traitements massifs ont eu lieu en Afrique et au Brésil dans le but d'endiguer la progression de la maladie, mais la réinfection rapide et élevée des populations locales reste un problème majeur (Parker & Allen 2011). Un autre problème est la découverte de certaines souches de *S. mansoni* résistantes au praziquantel (Ismail *et al.* 1994; Gryseels *et al.* 2001; Couto *et al.* 2011). Bien que ce phénomène ne semble pas pour l'instant très répandu sur le terrain (Greenberg 2013), il reste préoccupant et met en avant la nécessité de mieux comprendre par quels mécanismes le parasite arrive avec une telle efficacité à contourner les traitements. De plus, les stratégies vaccinales développées jusqu'à aujourd'hui ont toutes été soldées par des échecs. Certains vaccins candidats ont montré en laboratoire une protection contre certaines souches du parasite, mais se sont systématiquement

montrés inefficaces à grande échelle (McManus & Loukas 2008; Siddiqui *et al.* 2014). L'adaptabilité rapide du parasite aux différentes tentatives de traitements ou de préventions mérite une attention particulière.

1.3.2.3. Plasticité environnementale, de possibles Dauermodifikationen?

S. mansoni a la spécificité d'être une espèce très plastique, en particulier par sa capacité d'infection d'une grande diversité d'hôtes. Chez l'hôte intermédiaire du genre *Biomphalaria*, au moins 18 espèces, vivant dans des aires géographiques variées, sont connues pour permettre le développement larvaire du parasite (DeJong *et al.* 2001). Les espèces les plus notables sont *B. alexandrina* (Égypte), *B. pfeifferi* (Afrique tropicale, subtropicale et Moyen-Orient) et *B. glabrata* (Amérique du Sud et Caraïbes). Au niveau de l'hôte mammalien, une quarantaine d'espèces de rongeurs et de primates sont sensibles à *S. mansoni* (Martins 1957). De manière surprenante, ce parasite est capable d'infecter en laboratoire des hôtes vertébrés avec lesquels il n'est jamais en contact en milieu naturel : souris (*Mus musculus*), rat du Nile (*Arvicanthis niloticus*) ou hamsters (*Mesocricetus auratus*) (Martins 1957). Le large éventail d'hôtes, tant intermédiaires que définitifs, favorise la dispersion du parasite sur de longues distances et lui permet de facilement coloniser de nouveaux habitats. Un exemple de ce phénomène est l'introduction de *S. mansoni* en Amérique du Sud et aux Caraïbes au XVII^{ème} siècle lors de la traite négrière. Les parasites, originaires d'Afrique et transportés par les esclaves, ont trouvé un nouvel hôte chez *Biomphalaria glabrata*, une espèce uniquement présente au Nouveau Monde (Morgan *et al.* 2001) et ayant divergé de *Biomphalaria pfeifferi*, l'hôte intermédiaire principal en Afrique, depuis plusieurs millions d'années (DeJong *et al.* 2001). Il a été mis en évidence que des changements au niveau des souches de l'hôte intermédiaire ont des répercussions phénotypiques, et plus spécifiquement sur des traits de vie (sex-ratio, croissance, prévalence, intensité, etc.) du parasite en développement (Dias *et al.* 1988a; b; Lepesant *et al.* 2013; Theron *et al.* 2014). Des travaux effectués par l'unité de recherche UMR 5244 IHPE laissent penser que certains de ces phénotypes puissent être transgénérationnels. L'environnement direct du parasite étant le corps de son hôte, il n'est pas surprenant que des modifications du second aient un impact sur le premier. Quelques pistes indiquent que des mécanismes épigénétiques sont impliqués dans l'adaptation de *S. mansoni* à ses hôtes.

Un autre exemple de plasticité phénotypique induit par l'environnement a été décrit par Jansma *et al.* (1977). Alors qu'il travaillait sur l'hycanthone, une molécule antihelminthe brièvement utilisée pour traiter la schistosomiase, Jansma remarqua l'apparition d'une résistance métastable chez certaines souches de *S. mansoni*. La métastabilité signifie que la proportion d'individus résistants ou sensibles est hautement fluctuante d'une génération à l'autre, et ne correspond pas à ce que l'on attend d'une héritabilité mendélienne. Dans sa publication de 1977, il décrit trois méthodes pour induire cette résistance : en récupérant la progéniture de vers adultes après l'administration chez l'hôte rongeur d'une dose sous-curative d'hycanthone (3-60 mg/kg, la dose curative tuant >90% des vers adultes étant 80 mg/kg) 54 à 70 jours après l'infestation cercarienne (Type I), 27 à 29 jours après l'infestation quand les vers sont encore à un stade immature (Type II) et à partir d'hôtes infestés par des cercaires d'un seul sexe et, 2 à 58 semaines plus tard, par les cercaires du sexe opposé (Type III). Dans les trois cas, les parents sont susceptibles à la molécule thérapeutique alors qu'un fort pourcentage de la descendance survit à l'administration d'une dose curative d'hycanthone. La résistance a été observée comme métastable à différents degrés (en fonction du type d'induction utilisé) sur 10 à 21 générations. Les changements de la proportion d'individus résistants au sein de la population peuvent être drastiques, passant de 100% à moins d'un quart, et remontant ensuite à 90%, en l'espace d'une poignée de générations. Ainsi, l'héritabilité de ce phénotype ne semble pas suivre les règles classiques de la génétique mendélienne. Jansma a émis des hypothèses sur un possible phénomène héritable que l'on pourrait aujourd'hui associer à des modifications épigénétiques.

I.3.4. État de l'art des connaissances épigénétiques chez *S. mansoni*

L'épigénome de *S. mansoni* est en phase d'exploration. La présence de la méthylation de l'ADN est controversée. Elle serait globalement très faible et limitée à une fenêtre développementale très ciblée, mais il semblerait qu'elle joue un rôle au niveau de l'oviposition (Fantappié *et al.* 2001; Geyer *et al.* 2013; Raddatz *et al.* 2013). Les annotations du génome et du transcriptome ont montré une structure très conservée des histones et des enzymes les modifiant (Berriman *et al.* 2010; Protasio *et al.* 2012). Sur ce point, mon laboratoire d'accueil a mis au point la technique d'immunoprécipitation de chromatine suivi de séquençage (ChIP-Seq) en ciblant des modifications d'histones, sur miracidia, cercaires et vers adultes (Cosseau *et al.* 2009; Cosseau & Grunau 2011). Plusieurs études ont

récemment été consacrées à la caractérisation des ARNs non-codants chez *S. mansoni*. Elles ont dressé un inventaire des micros ARNs (miRNA) chez cet organisme à différents stades de développement (Simões *et al.* 2011) et identifiant éventuellement leurs cibles (de Souza Gomes *et al.* 2011). D'autres ont mis en évidence des différences d'expression des miRNA entre mâles et femelles adultes (Marco *et al.* 2013). On note de nombreuses publications sur ces ARNs non-codants chez *S. japonicum* (Cai *et al.* 2013), suggérant leurs rôles dans la maturation sexuelle des femelles chez ce parasite (Sun *et al.* 2014), leurs cibles possibles (Sun *et al.* 2014) et révélant leur présence circulante dans l'organisme de l'hôte définitif (Zhu 2015).

Un pan important de l'interaction de compatibilité/incompatibilité avec les hôtes mollusques est sous contrôle épigénétique. Bien que leur action exacte reste encore à éclaircir, les *SmPoMucs* (*Schistosoma mansoni* polymorphic mucins) sont des intermédiaires essentiels au succès d'infection de l'hôte intermédiaire. Cette famille de gènes est exprimée au moment de la pénétration dans le mollusque sous la forme de variants très divers d'un miracidium à un autre (Roger *et al.* 2008). Des différences génétiques existent entre les promoteurs de ces gènes, mais ne suffisent pas à expliquer la grande diversité d'expression observée. On retrouve par contre des structures chromatinienne très conservées et héritées de manière non mendélienne qui peuvent expliquer les différences d'expressions entre individus (Perrin *et al.* 2013; Fneich 2014). D'autres expériences avec la trichostatine A (TSA), un inhibiteur d'histones déacétylases, permettent d'observer un changement vers une structure euchromatique au niveau des gènes *SmPoMucs*. Cela se traduit par une forte augmentation de la diversité de l'expression de ces protéines chez les miracidia (contre seulement quelques variants sans l'ajout de TSA) (Cosseau *et al.* 2010).

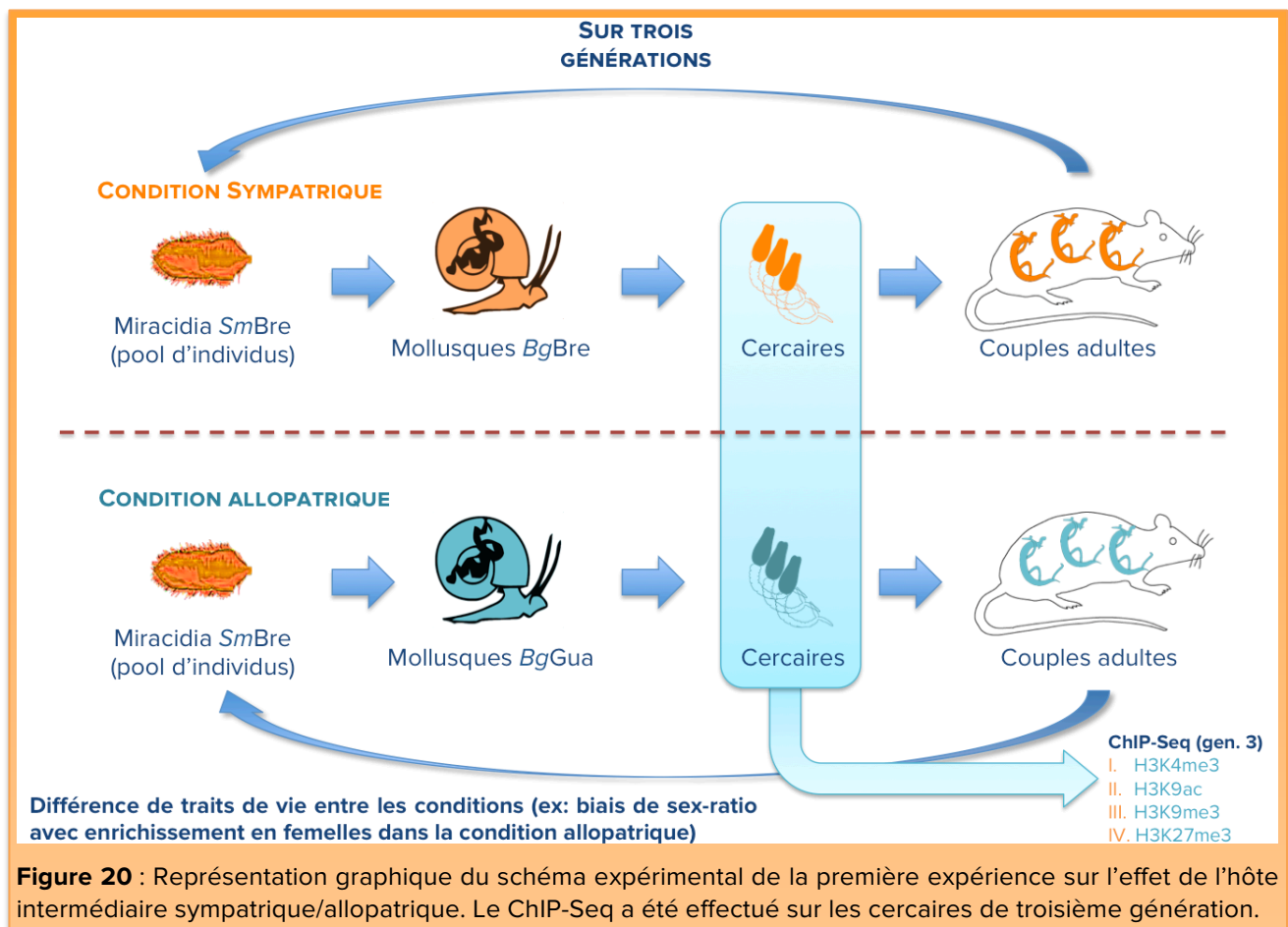
L'étude des mécanismes épigénétiques chez *S. mansoni* vise aussi le développement de nouvelles molécules thérapeutiques (Cabezas-Cruz *et al.* 2014). Par exemple, l'histone déacétylase 8 du parasite (*smHDAC8*) est essentielle à la survie des schistosomes. Sa structure étant notablement différente de celle des vertébrés, cela en fait un bon candidat pour des inhibiteurs ciblant le parasite, mais n'ayant pas d'effets sur son hôte (Marek *et al.* 2013).

I.3.5. Précédentes expériences et résultats

Avant mon arrivée en thèse, l'équipe de Christoph Grunau (laboratoire UMR 5244 IHPE) avait initié plusieurs expériences visant à observer si certains changements de conditions environnementales (écologiquement réalistes) pouvaient induire des phénotypes héréditaires chez *S. mansoni*, et leurs liens avec des changements de la structure chromatinienne. Ces travaux visaient à établir des profils de quelques modifications d'histones, à l'échelle du génome, sur plusieurs stades développementaux.

I.3.5.1. Impact de la souche d'hôte intermédiaire sur le développement parasitaire

Une première expérience (**Figure 20**) était destinée à observer l'impact de différentes souches d'hôtes intermédiaires (mollusques) sur l'épigénome de *S. mansoni*. Cette expérience a été réalisée dans le cadre de la thèse de Julie Lepesant (Lepesant 2012). Des miracidia d'un isola brésilien (*SmBre*) maintenus de manière consanguine depuis les années 1960 ont été utilisés pour infecter deux souches de *B. glabrata* avec des origines géographiques distinctes: une souche sympatrique du Brésil (*BgBre*) et une souche allopatrique de Guadeloupe (*BgGua*).



La souche *SmBre* a été sélectionnée à l'époque à cause de son apparente très faible diversité génétique, basée sur l'absence de polymorphisme sur quatorze marqueurs microsatellites neutres. Cette très faible diversité offre un contexte favorable à l'étude des modifications épigénétiques. C'est également une lignée pure concernant sa norme de réaction, c'est à dire que tous les individus de cette souche réagissent de la même manière aux changements environnementaux auxquels ils ont été soumis expérimentalement (ce qui n'est évidemment pas exhaustif). Des travaux ultérieurs sur cette souche montreront néanmoins qu'il existe une certaine variabilité génétique au sein de la population, certes très faible, mais tout de même non négligeable (Clément *et al.* 2013) (**Annexe 1**). Son développement chez les deux hôtes mollusques présente des différences de certains traits de vie, en particulier pour la prévalence (ratio d'hôtes infectés/sains), l'intensité (nombre moyen de parasites au sein d'un hôte) et le sex-ratio (proportion de cercaires mâles/femelles produite dans la population) (Lepesant *et al.* 2013; Theron *et al.* 2014). En condition allopatrique, la proportion de cercaires femelles est systématiquement plus élevée (66%) qu'en condition sympatrique (50%), et persiste sur plusieurs générations du parasite (Lepesant *et al.* 2013). Les cercaires émises par les deux souches de mollusques allopatriques et sympatriques ont été collectées et étudiées par immunoprécipitation de la chromatine suivie d'un séquençage massif de l'ADN précipité (ChIP-Seq) par la technologie Illumina *single-end 50 bp*. Cette approche à l'échelle du génome, permet l'identification de séquences d'ADN interagissant avec des modifications d'histones précises (ciblées par des anticorps spécifiques). Quatre modifications d'histones ont été étudiées. Deux sont généralement associées à une structure euchromatique (triméthylation de la lysine 4 sur l'histone H3 [H3K4me3] & acétylation de la lysine 4 sur l'histone H3 [H3K9Ac]) et deux autres à une structure hétérochromatique (triméthylation de la lysine 9 sur l'histone H3 [H3K9me3] & triméthylation de la lysine 27 sur l'histone H3 [H3K27me3]). Un contrôle négatif sans anticorps est aussi effectué. En parallèle, l'ARN messager des cercaires a été extrait, et séquençé par la technologie Illumina (RNA-Seq, *single-end 100bp*). L'objectif du RNA-Seq était d'identifier des gènes différentiellement exprimés entre les conditions allopatriques et sympatriques, et éventuellement les corrélérer à des modifications de la structure chromatinienne. Ces procédures ont été répétées sur trois générations du parasite. Au début de ma thèse, les résultats de ChIP-Seq de la troisième génération et de RNA-Seq des trois générations étaient disponibles.

À l'échelle du génome, une centaine de différences de modifications d'histones ont été identifiées. La majorité des variations a été observée avec H3K4me3 (voir **Figure 21** pour un exemple), une marque généralement présente sur des gènes compétents à la transcription, et H3K27me3 qui est associée à des régions d'hétérochromatine facultative et à la répression de la transcription (Berger 2007). Aucune corrélation directe entre l'intensité des marques d'histones et le niveau de transcription n'a été mise en évidence. Cette première approche a tout de même montrée que des changements de structure chromatinienne peuvent être associés à des environnements différents (dans ce cas ci, des hôtes différents). Ces différences de marques chromatiniennes ont été observées sur un seul stade développemental, avec un seul réplica biologique et à partir de pool de miracidia *smBRE*. Cela soulève trois questions majeures : **(I)** Ces différences sont-elles ciblées ou aléatoires **(II)** sont-elles transmises au stade développemental suivant et **(III)** dépendent-elles des différences génétiques existant entre les individus? Une de mes motivations à poursuivre ces travaux provient du biais de sex-ratio observé (Lepesant *et al.* 2013) dans les populations allopatriques. Il montre qu'il y a bien un effet de l'environnement sur le parasite, mais empêche de conclure sur l'origine des différences de structures chromatiniennes observées entre condition allopatrique et sympatrique. Sont-elles vraiment induites par l'environnement, ou s'agit-il de changements sexe-spécifiques que l'on observe à cause de l'enrichissement en femelle dans les mollusques *BgGua*?

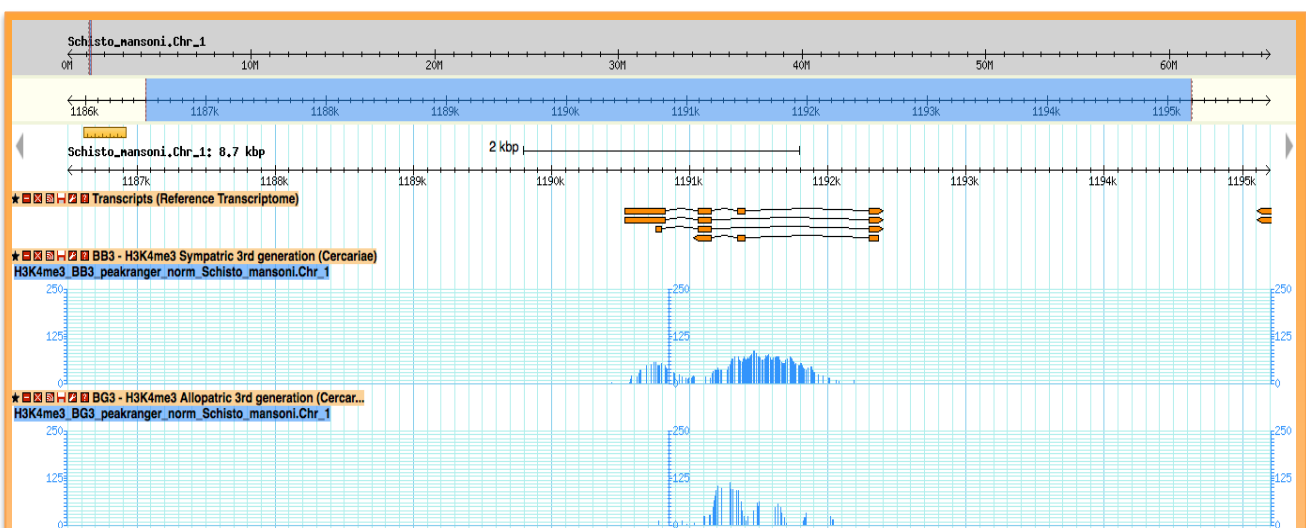


Figure 21 : Exemple de différence pour H3K4me3 entre les deux conditions expérimentales à une position génomique donnée. La piste « transcripts » illustre des données de RNA-Seq chez les cercariae, avec les exons en gras et les introns en ligne mince pour chaque transcrit. La deuxième piste correspond à la position de la marque H3K4me3 dans la condition sympatrique, et la troisième piste, la même marque dans la condition allopatrique. Intensité et forme des pics montrent un enrichissement différent de cette marque entre les deux conditions, ce qui pourrait s'expliquer biologiquement par des isoformes de transcrits différentes dans les deux conditions.

1.3.5.2. Résistance induite à un schistosomicide

Dans un deuxième temps, d'autres expériences effectuées au cours du doctorat de Julie Lepesant sous l'encadrement de Christoph Grunau ont permis d'initier l'étude du phénomène de résistance à l'antihelminthe hycanthone. Cette molécule, dont la structure et le mode d'action sont proches d'un autre schistosomicide (l'oxamniquine), a été brièvement utilisée dans les années 70 avant d'être abandonnée à cause de sa toxicité chez l'être humain (Galdino da Rocha Pitta 2013). Plusieurs publications ont signalé que l'exposition à des doses sous-curatives d'hycanthone (insuffisantes pour tuer la majorité des vers adultes) sous forme d'injection dans l'hôte vertébré murin induisait une résistance à cette molécule chez la progéniture des parasites exposés (Jansma *et al.* 1977; Brindley & Sher 1987; Brindley *et al.* 1989). La résistance était décrite comme métastable, c'est-à-dire ne suivant pas les règles de l'hérédité mendélienne, et était extrêmement fluctuante d'une génération à l'autre. Deux autres méthodes d'induction de résistance relativement similaires, menant également à une résistance métastable, ont été documentées (Jansma *et al.* 1977). Ces travaux n'ont cependant pas pu être reproduits par toutes les équipes travaillant sur ce phénomène (Dias & Oliver 1986; Cioli *et al.* 1992), et il est possible que seules certaines souches (ou génotypes) du parasite soient susceptibles à cette induction de résistance à l'hycanthone.

La métastabilité du phénotype de résistance, ainsi que sa capacité d'induction par de faibles doses d'hycanthone, en faisait un bon candidat pour l'étude de modifications épigénétiques. Pour explorer cette possibilité, nous avons travaillé avec une souche du laboratoire pour laquelle nous avons observé cette induction de résistance à l'hycanthone. Des cercaires issues d'une infestation mono-miracidiale (un seul miracidia par mollusque, donc toutes les cercaires émises sont génétiquement identiques) ont été utilisées pour infecter cinq souris. Les souris ont été traitées avec une dose sous-curative d'hycanthone de 60 mg/kg, telle que décrite dans Jansma *et al.* (1977) et les vers survivants ont été collectés. Cette expérience a été faite à partir des cercaires issues de cinq mollusques infectés séparément (donc cinq génotypes de *S. mansoni*). Des analyses de ChIP-Seq, ciblant les mêmes quatre marques décrites dans la précédente expérience sur l'effet de la souche d'hôte intermédiaire (H3K4me3, H3K9Ac, H3K9me3, H3K27me3), ont été réalisées sur les vers survivants au traitement à l'hycanthone, ainsi que sur des vers adultes contrôles. Ces résultats ont été analysés et interprétés dans le cadre de mon doctorat.

I.4. LE CORAIL TROPICAL *POCILLOPORA DAMICORNIS*

I.4.1 Description du modèle

I.4.1.1. Caractéristiques biologiques et cycle de vie

Pocillopora damicornis, aussi appelé corail chou-fleur ou encore corail verruqueux (**Figure 22 A**), est une espèce répandue de corail scléractiniaire (corail dur) du phylum des *cnidaria*. Il forme des colonies sessiles n'excédant généralement pas 30 cm de hauteur. Les colonies sont composées d'individus génétiquement identiques, les polypes, et croissent par multiplication asexuée de ces derniers. Comme plusieurs autres espèces du même genre, il est caractérisé par la présence de corallites (niches squelettiques dans lesquelles sont logés les polypes) de tailles constantes (environ 3 mm de diamètres) et dont la forme rappelle des verrues, répartis de manière irrégulière sur les branches. *P. damicornis* est un organisme diploblastique (ne possédant que deux feuilletts embryonnaires : endoderme et ectoderme) et bénéficie d'une morphologie relativement simple (**Figure 22 B**).

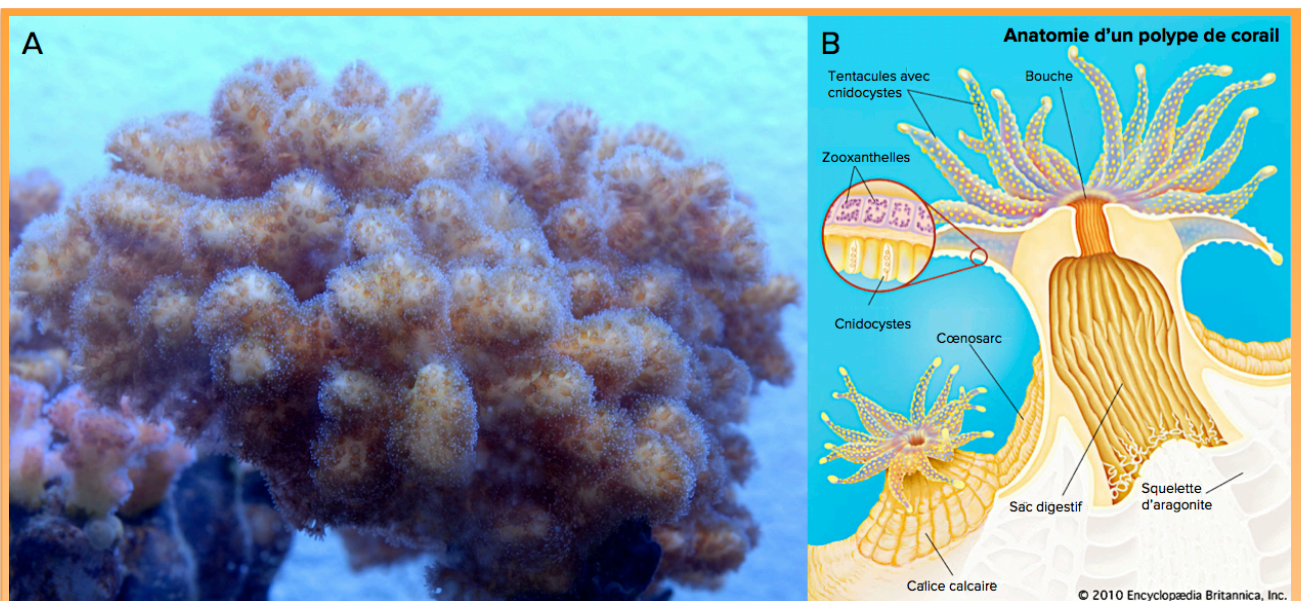


Figure 22 : (A) Colonie de *Pocillopora damicornis*. Photographie de Kelly Brener-Raffalli. **(B)** Morphologie d'un polype corallien au sein d'une colonie. Figure tirée de britannica.com.

On y retrouve un pied contenant une cavité digestive, surmontée d'un disque oral pourvu de tentacules (présentes en multiples de 6). Les tentacules sont couverts de cnidocytes, des cellules éjectant, lors d'un contact physique, un filament urticant à la manière d'un harpon. Ils servent à chasser des proies planctoniques. Durant sa croissance, la colonie génère un exosquelette rigide d'aragonite (coenostéum), une forme de carbonate

de calcium. Les polypes sont reliés entre eux sur le squelette par le cœnosarc. La croissance fluctue selon les origines géographiques des souches, allant de 9 mm par an dans certains récifs australiens et jusqu'à 54 mm par an à Panama (Hughes *et al.* 2015).

C'est une espèce hermaphrodite simultanée de type « *brooder* » (fécondation interne et libération d'une larve dite planulæ déjà formée). Elle peut exceptionnellement avoir des phases « *broadcaster* » avec fécondation externe après libération des gamètes mâles et femelles dans la colonne d'eau (**Figure 23 A**) (Ward 1992). Les libérations de planulæ sont mensuelles et peuvent avoir lieu toute l'année, en fonction de la température de l'eau (Fan *et al.* 2002). La planulæ connaît une courte phase pélagique avant de se fixer sur un substrat pour se développer en polype, puis former une colonie. Une grande diversité de modes de multiplication asexuée est connue chez *P. damicornis* (**Figure 23 B**).

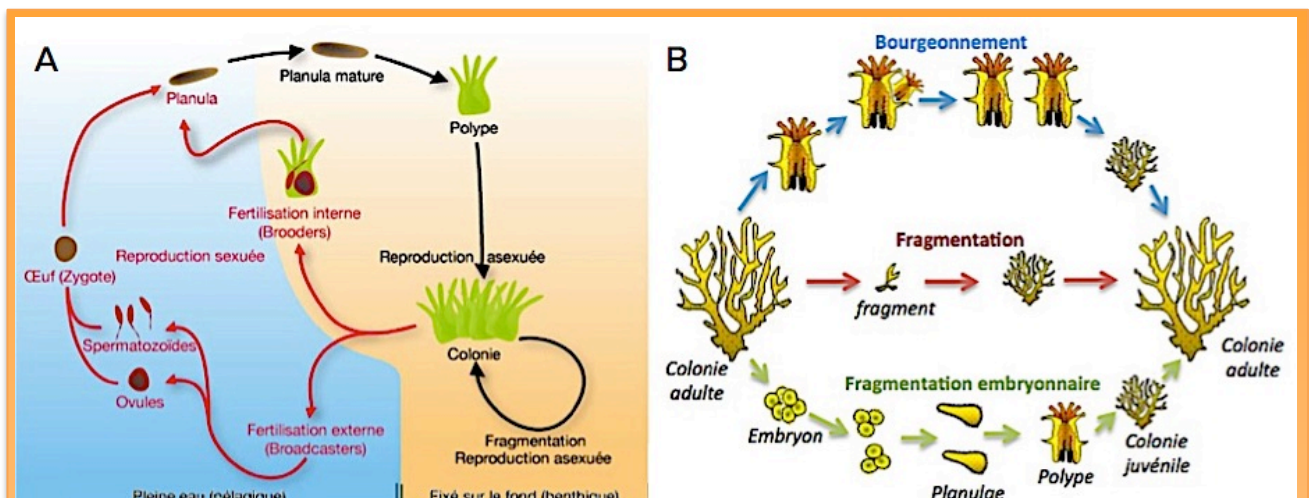


Figure 23 : Modes de reproduction connus chez *P. damicornis*. **(A)** On peut distinguer deux grands types de reproduction sexuée : par fertilisation externe (*broadcasters*) et interne (*brooders*). Dans le premier cas, les gamètes mâles et femelles sont libérés dans la colonne où la fécondation a lieu. L'œuf fécondé se développera en larve planula. Dans le second, seules les gamètes mâles sont expulsés et la fécondation est interne. Des planulæ seront ensuite libérées par les polypes. **(B)** Trois types de reproduction asexuée : *bourgeonnement* (un nouveau polype se développe à partir des tissus d'un autre), *fragmentation* (un fragment issu d'un bris accidentel d'une colonie se fixe sur un substrat et génère une nouvelle colonie) et *fragmentation embryonnaire* (un embryon formé par fertilisation externe se scinde en deux, qui se développeront en deux planulæ, puis deux polypes et finalement deux colonies).

Au sein d'une colonie, la croissance se fait par bourgeonnement, qui est la génération d'un nouveau polype par division d'un polype existant (bourgeonnement intra-tentaculaire) ou développement sur des tissus adjacents (bourgeonnement extra-tentaculaire). Le polype nouvellement formé peut aussi se détacher sous son propre poids et générer une colonie fille (Kramarsky-Winter *et al.* 1997). Un fragment brisé d'une colonie (souvent par l'action des vagues) peut se fixer sur un substrat et croître en une nouvelle

colonie (Harrison 2011). Les embryons sont également sensibles aux perturbations hydrodynamiques et peuvent se fragmenter sous l'effet de vagues. Les fragments d'embryons peuvent ensuite se développer en planulæ et continuer leur cycle de vie normalement (Heyward & Negri 2012). Une expulsion de polypes est susceptible de se produire sous l'effet d'un stress environnemental ou d'une pression de prédation. Un polype d'une colonie peut se décrocher et se laisser dériver pour se fixer plus loin. La colonie mère ne survit et ne guérit habituellement pas (Sammarco 1982). Finalement, chez quelques *brooders*, des œufs non-fécondés peuvent se développer en planulæ parthénogénétique, avant d'être libérées dans la colonne d'eau (Harrison 2011).

P. damicornis est un corail hermatypique, ce qui signifie qu'il entretient une relation symbiotique mutualiste avec des zooxanthelles (dinoflagellés photosynthétiques du genre *Symbiodinium*) présentes à l'intérieur des tissus. L'endosymbionte génère par autotrophie la majeure partie de l'énergie requise pour le processus de calcification, en échange du dioxyde de carbone et des composés azotés produits par le polype. Plusieurs clades de *Symbiodinium* peuvent coloniser *P. damicornis*, chacun d'entre eux ayant des spécificités physiologiques (Cooper *et al.* 2011). Par exemple, les *Symbiodinium* du clade C sont de bons producteurs photosynthétiques, mais sont sensibles au stress, alors que ceux du clade D sont moins efficaces, mais plus résistants au stress environnemental.

I.4.1.2. Distribution géographique

L'espèce est native des régions tropicales et subtropicales des océans Indien et Pacifique (**Figure 24**). Elle est abondante le long des côtes de l'Afrique de l'Est et de la Mer Rouge, dans les mers du Japon et de l'Indonésie, en Polynésie, Australie, Golfe d'Oman, et sur la partie Ouest de l'Amérique Centrale (IUCN 2014). L'espèce est présente, mais rare, au niveau des récifs de l'Équateur, du Mexique et des îles Revillagigedo, et a presque complètement disparue des îles Cañao, de Panama et de l'archipel des Galapagos suite à l'évènement El Niño de 1982/1983 (IUCN 2014). Elle est principalement distribuée à faible profondeur. Très courante entre 1 et 15 m, on la retrouve rarement à plus de 18-20 m (Spalding *et al.* 2001). Ce corail s'accommode relativement bien à la sédimentation et une faible salinité (dans les estuaires ou mangroves par exemple), mais est sensible à l'hydrodynamie. L'espèce a besoin d'agitation, mais pas excessive (Spalding *et al.* 2001).

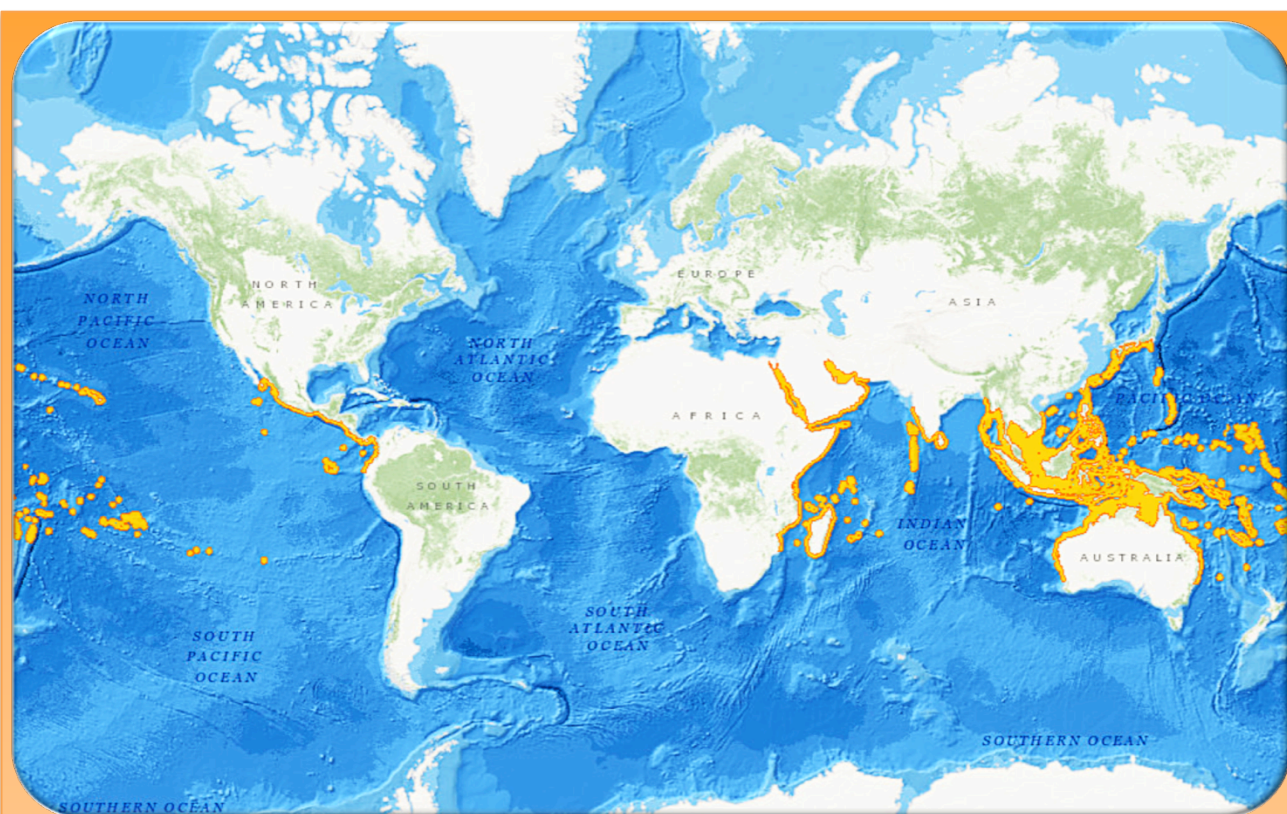


Figure 24 : Distribution des récifs de *Pocillopora damicornis* (en jaune). Figure tirée de IUCN (2014).

I.4.1.3. Importance écologique et économique

P. damicornis est un important architecte des récifs coralliens. Étant une espèce compétitrice et avec un rapide taux de croissance, elle joue un rôle crucial dans l'écologie de ces environnements, entre autres en fournissant un habitat complexe et diversifié. Les récifs coralliens sont considérés comme étant des écosystèmes clefs caractérisés par une très forte biodiversité, plus importante que celle de la forêt amazonienne et estimée à plus de 950 000 espèces (Hughes *et al.* 2002). Ils ont une grande importance au niveau des industries de la pêche et du tourisme, et offrent une protection contre l'érosion des côtes (Hughes *et al.* 2002). Nombre d'entre eux sont menacés par le réchauffement climatique et d'autres facteurs anthropiques (Hoegh-Guldberg *et al.* 2007; Hoegh-Guldberg 2011).

I.4.1.4 Plasticité environnementale : thermotolérance

Comme la plupart des coraux hermatypiques, *P. damicornis* est sensible aux perturbations climatiques. Si la température de l'eau dépasse un seuil de tolérance, spécifique à chaque espèce, cela induit un phénomène de blanchissement (**Figure 25**). Il s'agit de l'expulsion des zooxanthelles endosymbiotiques par les polypes. Sans sa source

d'énergie principale, l'hôte cnidaire peut survivre de quelques jours à quelques semaines par hétérotrophie, en chassant des proies planctoniques. Cependant, si la température de l'eau ne redescend pas sous le seuil de tolérance, permettant la recolonisation des zooxanthelles, la colonie corallienne finira par mourir. En fonction de leur répartition géographique, tous les récifs de *P. damicornis* n'ont pas les mêmes températures optimales de croissance et ne sont pas soumis aux mêmes niveaux de fluctuations annuelles de température (Palumbi *et al.* 2014), ce qui montre une certaine plasticité de l'espèce.

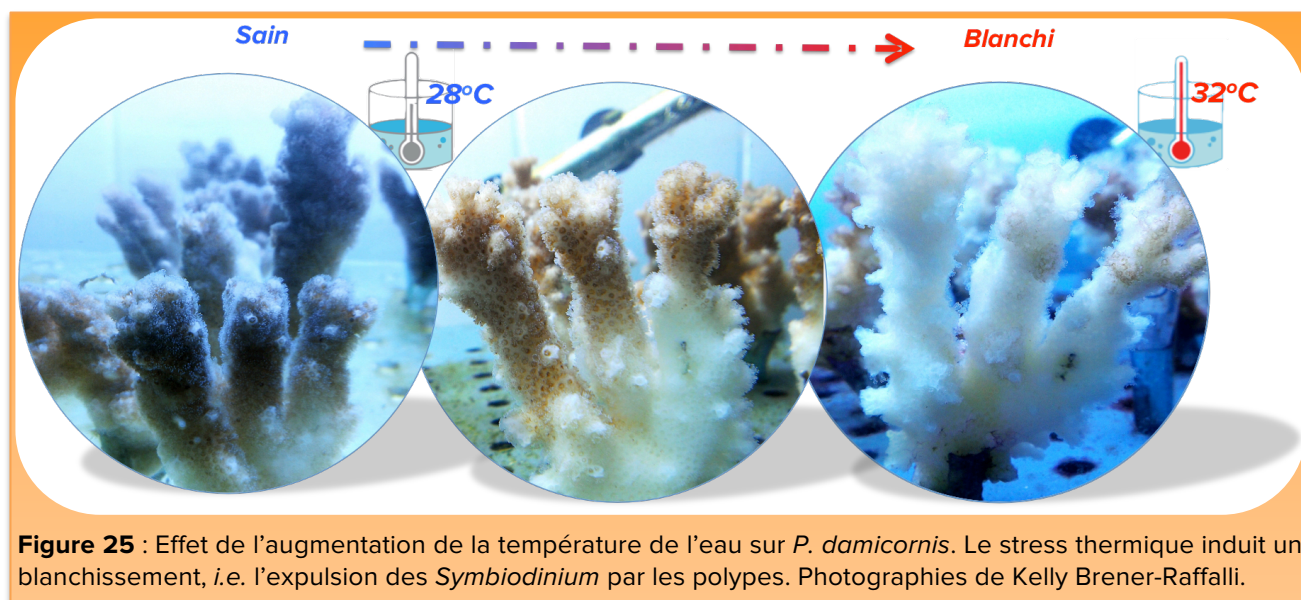
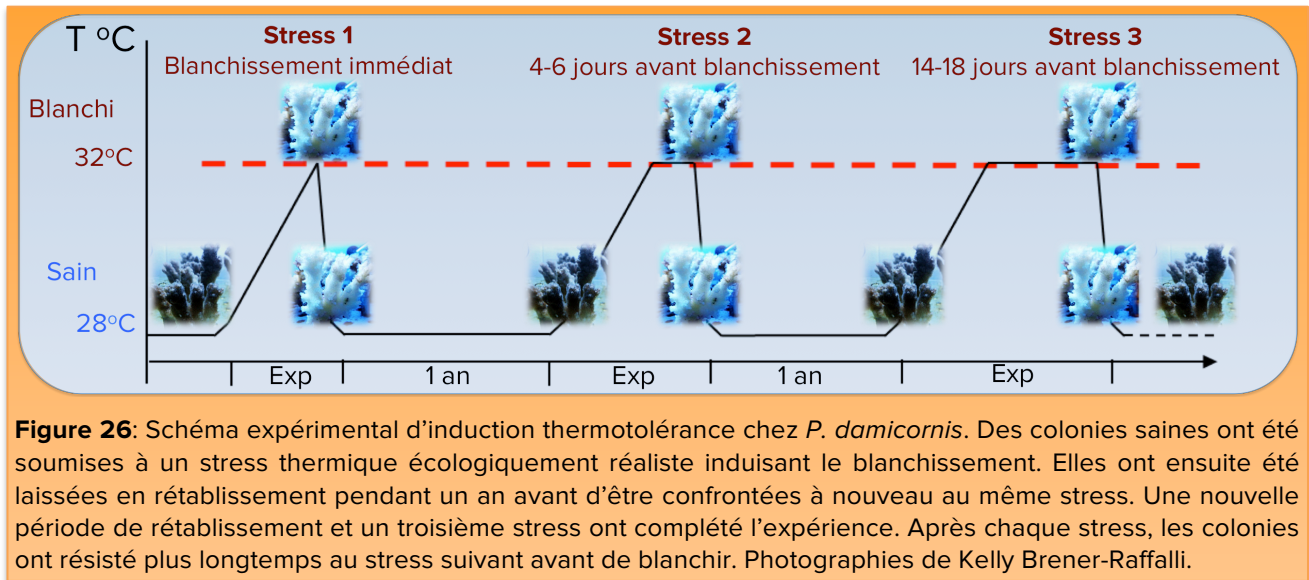


Figure 25 : Effet de l'augmentation de la température de l'eau sur *P. damicornis*. Le stress thermique induit un blanchissement, i.e. l'expulsion des *Symbiodinium* par les polypes. Photographies de Kelly Brener-Raffalli.

Des travaux effectués par Jérémie Vidal-Dupiol au sein du laboratoire UMR 5244 IHPE (Vidal-Dupiol 2011; Vidal-Dupiol *et al.* 2011) ont mis en évidence une plasticité dans la résistance au stress thermique chez *P. damicornis* (**Figure 26**). Des colonies filles issues d'une colonie mère (donc toutes génétiquement identiques) ont été soumises à un stress thermique écologiquement réaliste induisant un blanchissement immédiat une fois le seuil de tolérance dépassé. Les colonies blanchies ont ensuite été mises en rétablissement pendant une année, permettant la recolonisation par *Symbiodinium*, avant de subir un second stress thermique identique. Cette fois ci, les coraux ont résisté 4 à 6 jours avant de blanchir. Suivant une nouvelle période de rétablissement d'un an, un troisième et dernier stress thermique a été imposé aux colonies. Le blanchissement n'est alors apparu que 14 à 18 jours après le début du stress. Dans la littérature, une expérience de transplantation sur le corail *Acropora hyacinthus* entre deux localités aux conditions climatiques différentes a montré que les colonies s'adaptent à leurs nouvelles conditions environnementales en quelques années et deviennent moins sensibles au blanchissement (Palumbi *et al.* 2014).



1.4.2. État de l'art des connaissances épigénétiques chez *P. damicornis*

La littérature sur l'épigénome de *P. damicornis*, ou des coraux scleractiniaires en général est quasi inexistante. Au début de ma thèse, des résultats de transcriptomique montraient que des orthologues d'enzymes interagissant avec la chromatine existent chez cette espèce (Vidal-Dupiol, communication personnelle), que ce soit au niveau de la méthylation des cytosines ou des modifications d'histones. Le génome de *P. damicornis* a été séquencé, assemblé et annoté par notre laboratoire durant la période couvrant mon doctorat. Il est de taille moyenne, environ 325 mb. Pour *Symbiodinium*, son génome séquencé par Shoguchi *et al.* (2013) est assez gros (environ 3 000 mb), comporte un fort taux de 5-méthylcytosone et ne contient ni histones, ni nucléosomes (Hackett *et al.* 2005). Un des défis relié à l'étude de l'épigénome du corail consiste à séparer le matériel biologique du cnidaire de celui de *Symbiodinium*. À cause de la différence de taille entre les génomes de l'endosymbionte et de l'hôte, une contamination par un seul *Symbiodinium* apporte autant d'ADN que dix cellules de *P. damicornis*. Ma contribution dans ce projet sera de mettre au point une technique d'isolation des noyaux et de la chromatine de *P. damicornis*. Selon l'efficacité de cette technique, une approche d'étude du méthylome par séquençage bisulfite, ou d'analyse des modifications d'histones par ChIP-Seq (ou les deux) sera effectuée. Tel que mentionné précédemment, les dinoflagellés (dont fait partie *Symbiodinium*) sont les seuls eucaryotes connus dépourvus d'histones (Rizzo 2003). Cette caractéristique offre un aspect pratique dans l'éventualité où l'isolation des noyaux de *P. damicornis* ne soit pas concluante, et que seule une approche ChIP-Seq soit décidée.

I.5. OBJECTIFS DE LA THÈSE

I.5.1. Hypothèses et objectifs généraux

S'il commence à être évident qu'il existe des liens entre stimuli environnementaux, plasticité phénotypique et modifications de la structure chromatinienne, nous sommes encore loin d'appréhender comment ces liens sont établis au niveau moléculaire. Un second problème est d'être capable de bien mesurer les implications respectives des codes génétiques et épigénétiques, les deux étant étroitement liés. Mon but au commencement de cette thèse était de mieux explorer les liens entre des phénotypes induits par l'environnement (*Dauermodifikationen*) et des modifications de la structure chromatinienne. Je pouvais voir plusieurs hypothèses pouvant expliquer mes observations et celles de la littérature sur les *Dauermodifikationen*. **(I)** On peut premièrement supposer que l'environnement a un effet ciblé sur un caractère donné, et qui s'expliquera par une modification bien définie de la structure chromatinienne. Pour donner une illustration (fictive) très simple, un changement toujours identique d'une condition environnementale ira activer/inactiver l'expression d'un ou de plusieurs gènes (toujours les mêmes) et produira un phénotype adaptatif. Il est alors intéressant de regarder au sein d'une population si tous les individus répondent de la même manière à ce stimulus, peu importe leur génotype, et sinon quel est l'impact de ce dernier (en supposant, pour l'exemple, que le ou les gènes de réponse ne varient pas au sein de la population, mais que ce soit leur seule expression ou répression qui soit en jeu). **(II)** On peut aussi imaginer que l'environnement provoque des épimutations aléatoires, générant des phénotypes variés, et seuls les plus adaptatifs seraient sélectionnés. **(III)** À l'instar des mutations génétiques, il est aussi envisageable que ces épimutations aléatoires surviennent en permanence et que l'environnement ne soit pas inducteur, mais seulement agent de sélection sur les phénotypes provoqués par ces épimutations. **(IV)** Peut-être que ces épimutations aléatoires arrivent en temps normal, mais sont accrues par les stress environnementaux. **(V)** Il pourrait aussi s'agir de modifications d'ordre génétique qui n'auraient pas été détectées, et qui auraient éventuellement des effets en *cis* sur la structure chromatinienne. La question de l'héritabilité de ces *Dauermodifikationen* se pose également. S'il s'agit de changement de la structure chromatinienne, dans quelle mesure sont-ils transmissibles à une descendance somatique ou méiotique?

J'ai résumé toutes les interrogations soulevées par ces hypothèses sous la forme de deux questions principales, deux objectifs scientifiques, qui constitueront le socle de mon travail de recherche lors de cette thèse :

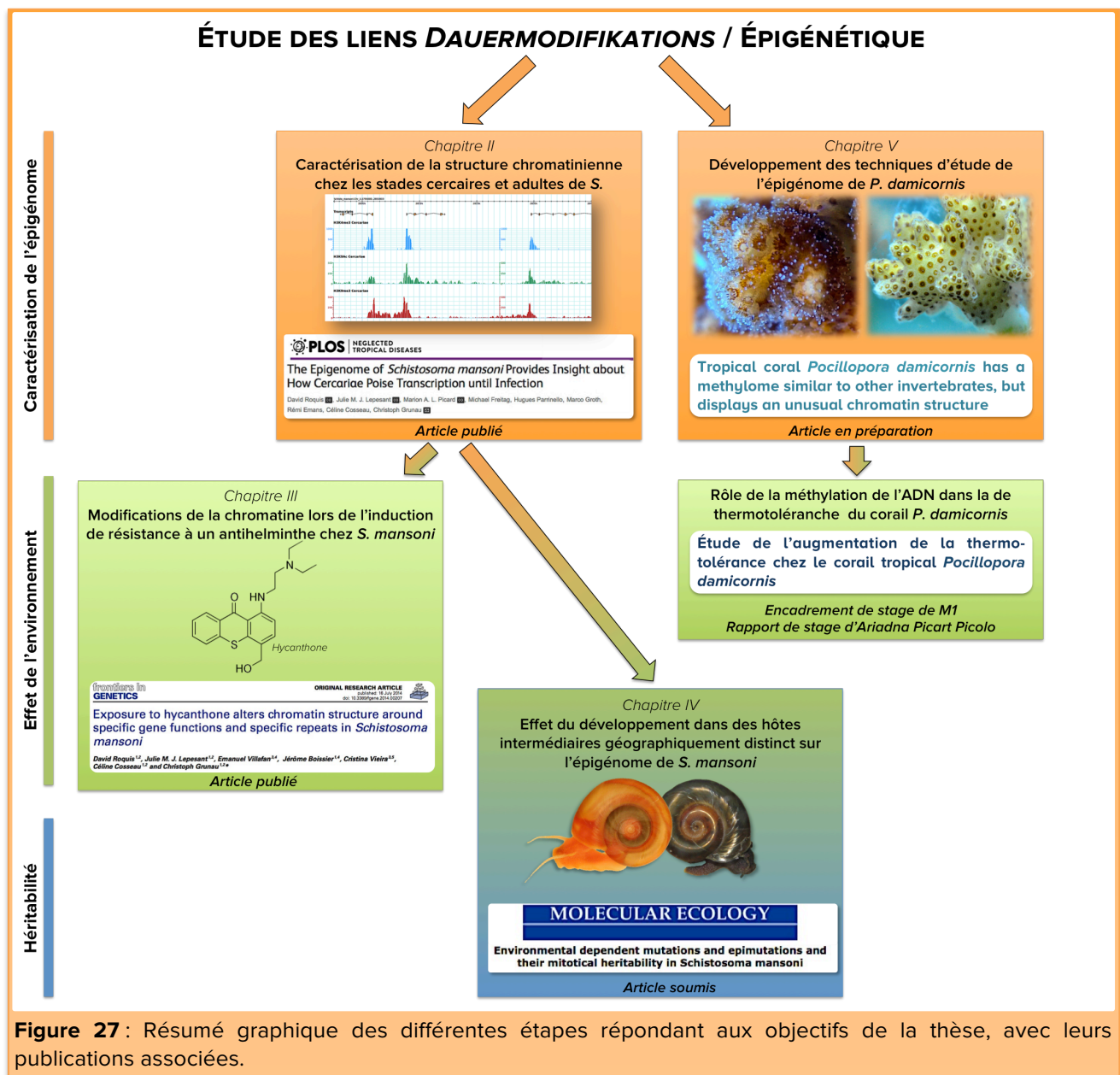
- 1. Les changements de structure chromatinienne observés entre des individus évoluant sous des conditions environnementales différentes sont-ils aléatoires, ciblés, ou les deux?**
- 2. Les modifications de la structure chromatinienne induites par un environnement changeant sont elles transmissibles mitotiquement/asexuellement et/ou au travers de la reproduction sexuée (méïotiquement)?**

Pour répondre à ces deux questions, j'ai mis au point des expériences afin de tester sans *a priori* les différentes hypothèses présentées plus haut, en explorant génomes et épigénomes sous l'influence de changements environnementaux. J'ai choisi deux modèles respectant les caractéristiques que j'ai définies pour l'étude des liens entre modifications épigénétiques et *Dauermodifikationen*. Leurs caractéristiques principales sont la présence d'une période de multiplication clonale lors du cycle de vie, pour garder un fond génétique constant, et des espèces en interactions (mutualisme pour le *P. damicornis* et parasitisme pour *S. mansoni*) car pour ces espèces, les pressions de sélection imposées par la coévolution avec leurs partenaires sont fortes et les phénomènes évolutifs sont rapides.

Ces deux modèles présentent des avantages et inconvénients dans le cadre des objectifs scientifiques de ma thèse. De nombreux outils moléculaires et informations sur le génome et transcriptomes sont disponibles pour *S. mansoni*, mais seule une petite fenêtre de son cycle de vie génère de la clonalité, et limite les possibilités d'études des *Dauermodifikationen* à cette partie de son cycle de vie. *P. damicornis* est beaucoup plus facile d'accès sur cet aspect, car on peut bouturer des colonies pour en obtenir de nouvelles, qui seront toutes génétiquement identiques. On peut ainsi soumettre un grand nombre de clones à des conditions environnementales variées pour en observer l'impact sur l'épigénome. Toutefois, tous les outils d'étude des modifications épigénétiques sont à mettre au point sur ce modèle, et aucune information sur son épigénome n'a été publiée à ce jour. Si on se réfère à la théorie de l'héritabilité inclusive (**Figure 6**, page 24), ces deux modèles ont aussi l'avantage de ne probablement pas avoir (ou très peu) de facteurs

d'hérabilité sociales ou culturelles pouvant jouer un rôle dans les *Dauermodifikationen* étudiées. Il est donc logique de se concentrer sur la dimension épigénétique. Dans le même ordre d'idée, aucune communauté bactérienne associée à *S. mansoni* n'a été trouvée (Eve Toulza, communication personnelle), mais il sera essentiel dans le cas *P. damicornis* de regarder au niveau de son holobionte si des variations sont détectées, en particulier dans la composition des clades de *Symbiodinium*.

Les deux grands objectifs se découpent en sous-objectifs spécifiques à chaque modèle (détaillés dans les sections suivantes et visibles sur la **Figure 27**). Une publication dans une revue scientifique est prévue pour chaque objectif. Deux articles ont déjà été publiés, un troisième est soumis et un quatrième est en préparation.



I.5.2. Objectifs spécifiques à *Schistosoma mansoni*

Puisque les techniques d'étude de l'épigénome de *S. mansoni* sont déjà au point, il a été décidé de découper l'étude des *Dauermodifikationen* chez ce parasite en trois parties.

I) il est nécessaire d'avoir une meilleure vision d'ensemble du fonctionnement de l'épigénome, pour mieux cerner l'influence que l'environnement peut avoir dessus. La méthylation de l'ADN étant très limitée chez le schistosome (Geyer *et al.* 2011), l'étude de la distribution de six modifications d'histones, à l'échelle du génome sur les stades développementaux cercaires et adultes, a été ciblée.

II) Les travaux de Jansma *et al.* (1977) sur l'induction de résistance à l'antihelminthe hycanthone offrent une première opportunité d'étudier une *Dauermodifikation* chez *S. mansoni*. Une répétition de son induction de type I a été faite par mon laboratoire d'accueil avant mon arrivée, générant des données de ChIP-Seq que je me charge d'analyser.

III) La seconde *Dauermodifikation* identifiée chez *S. mansoni* résulte en l'altération de traits de vie lors du développement chez deux hôtes intermédiaires d'origines géographiques distinctes. À partir de précédents travaux de mon laboratoire d'accueil, il transparaît qu'il y a bien un effet qui pourrait être transgénérationnel. Des expériences additionnelles sont essentielles pour mieux cerner l'effet exact sur l'épigénome, et de quelle manière il peut être héritable. Une des limites intrinsèque du modèle *S. mansoni* est l'impossibilité d'obtenir des miracidia clonaux pour infecter les différents mollusques, et ainsi avoir un génotype du parasite constant. Il est possible de contourner ce problème en effectuant des greffes de sporocystes. Il s'agit d'une technique microchirurgicale permettant, à partir d'un mollusque donneur infecté, de transférer des sporocystes secondaires (Sp2) dans une trentaine de mollusques receveurs (Jourdane & Theron 1980; Jourdane 1984). Les Sp2 retournent alors à l'état de sporocystes primaires (Sp1) et effectuent le reste du cycle intramolluscal normalement. Bien qu'écologiquement non réaliste, cette méthode est la seule permettant de maintenir un génotype constant entre plusieurs conditions et avec des réplicats. La structure chromatinienne des cercaires émises par les mollusques greffés sera analysée par ChIP-Seq. Des vers adultes obtenus par infection de souris à partir des mêmes cercaires serviront, également par ChIP-Seq, à déterminer si des différences observées entre des cercaires issues de mollusques différents sont transmissibles somatiquement au stade développemental suivant.

I.6.3. Objectifs spécifiques à *Pocillopora damicornis*

Mon travail sur les *Dauermodifikationen* chez le corail *P. damicornis* s'inscrit dans un projet à grande échelle incluant l'expérience de thermotolérance (**Figure 26**, page 60). Afin de cerner les mécanismes en jeu dans les adaptations rapides du corail à l'augmentation de température, des échantillons avant et après chacun des trois stress seront analysés sur le plan génomique (recherche de mutations somatiques), transcriptomiques, épigénomiques, et sur la composition en clades de *Symbiodinium* dans les tissus.

Ma responsabilité dans ce projet sera de mettre au point les outils moléculaires pour analyser la structure chromatinienne de *P. damicornis*. Mon premier objectif est de développer une technique permettant d'isoler les noyaux des cellules de *P. damicornis*, et d'éviter toute contamination avec du matériel biologique de *Symbiodinium*. En fonction des résultats, l'étude de l'épigénome s'effectuera soit en ciblant les modifications d'histones par ChIP-Seq et/ou la méthylation des cytosines par séquençage bisulfite (BS-Seq). Une fois ces techniques optimisées, je présenterai une première description de l'épigénome de *P. damicornis*.

Une première caractérisation des différences entre les épigénomes des coraux thermotolérants et contrôles (jamais stressés) a été effectuée par une étudiante de Master I que j'ai encadré, Ariadna Picart-Piccolo, et les résultats de son travail sont disponible dans son rapport de stage « *Étude de l'augmentation de la thermotolérance chez le corail tropical Pocillopora damicornis* ». J'ai fait le choix de ne pas présenter les détails de ses analyses dans ce manuscrit, car elles sont encore à un stade très préliminaire.

CHAPITRE II

CARACTÉRISATION DE LA STRUCTURE CHROMATINIENNE DE S. MANSONI

II.1. MISE EN CONTEXTE

Au commencement de ma thèse, l'épigénome de *S. mansoni* était toujours en phase exploratoire. La méthylation de l'ADN étant très rare chez cette espèce (Fantappiè *et al.* 2001; Geyer *et al.* 2011), mon laboratoire d'accueil a préféré se concentrer sur un autre porteur d'information épigénétique : les modifications d'histones. La mise au point de la technique d'immunoprécipitation de chromatine suivi de séquençage massif (ChIP-Seq) a été effectuée sur les trois principaux stades développementaux du parasite (miracidium, cercaire, adultes) par mes deux encadrants, Céline Cosseau et Christoph Grunau, peu avant mon arrivée au laboratoire (Cosseau *et al.* 2009; Cosseau & Grunau 2011). Le développement de la méthode fut loin d'être trivial, à cause de la petite taille (et donc de la faible quantité de matériel biologique) de l'organisme étudié, et de la difficulté d'utiliser les anticorps commerciaux sur ce modèle. À titre d'exemple, sur 37 anticorps testés dans le pipeline de validation présenté dans Cosseau & Grunau (2011), seuls six ont passé toutes les étapes nécessaires pour être utilisés en ChIP-Seq sur *S. mansoni*. Julie Lepesant a fourni durant sa thèse un travail expérimental remarquable en produisant, à l'échelle du génome, les données pour quatre modifications d'histones sur les stades cercaires et adultes (Lepesant 2012). Elle n'a cependant pas été en mesure d'analyser ces données, ce dont je me suis chargé à la fin de mon Master et durant la première année de ma thèse. Pour ce faire, j'ai développé un pipeline d'analyse bioinformatique mis en ligne sur notre serveur Galaxy (bioinfo.univ-perp.fr) (Goecks *et al.* 2010).

L'analyse de ces données était essentielle pour obtenir une meilleure compréhension des caractéristiques de l'épigénome de *S. mansoni*, avant de pouvoir me lancer dans un travail comparatif entre des conditions environnementales différentes. Dans ce chapitre, je me suis attardé à comprendre comment les quatre marques testées par Julie Lepesant se distribuaient sur le génome, et comment elles se comportaient par rapport à ce qui était connu chez d'autres organismes. Son choix des marques était particulièrement approprié, car il permettait de couvrir différents états chromatinien. En effet, deux marques (triméthylation de la lysine 4 sur l'histone 3 [H3K4me3] et acétylation de la lysine 9 de l'histone 3 [H3K9ac]) sont plutôt associées à des structures chromatinien permissives à l'expression génique, alors que les deux autres sont plutôt associées à une forme condensée et répressive de la transcription (triméthylation de la lysine 9 de l'histone 3

[H3K9me3] et triméthylation de la lysine 27 de l'histone 3 [H3K27me3]). Tel que je l'ai présenté dans l'introduction générale, la notion d'épigénétique est étroitement liée au développement. Chez *S. mansoni*, un seul génome permet de générer des phénotypes drastiquement distincts entre les différents stades (**Figure 18**, page 45) et il me semble rationnel que ces changements soient associés à des bouleversements de la structure chromatinienne. Un autre aspect de ma contribution dans ce travail aura donc été de caractériser ces différences chromatinienes entre cercaires et adultes, et de produire des données supplémentaires pour les marques H3K4me3 et H3K27me3 sur des schistosomules, un stade développemental intermédiaire des deux autres. J'ai aussi réalisé des expériences d'immunomarquage pour identifier l'impact transcriptionnel des modifications de la structure chromatinienne entre cercaires et schistosomules.

II.2. PUBLICATION ASSOCIÉE

Résultats et conclusions de ces travaux ont été publiés dans l'article « *The epigenome of Schistosoma mansoni provides insight about how cercariae poise transcription until infection* », en août 2015 dans la revue **PLoS Neglected Tropical Diseases**. La *Supplementary Figure 1* et les *Supplementary Tables 1* et *2* de l'article sont présentés à sa suite. Les *Supplementary Tables 3* et *4* sont accessibles en ligne à l'adresse suivante : <http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0003853>.

RESEARCH ARTICLE

The Epigenome of *Schistosoma mansoni* Provides Insight about How Cercariae Poise Transcription until Infection

David Roquis^{1,2}, Julie M. J. Lepesant^{1,2}, Marion A. L. Picard^{1,2}, Michael Freitag³, Hugues Parrinello⁴, Marco Groth⁵, Rémi Emans^{1,2}, Céline Cosseau^{1,2}, Christoph Grunau^{1,2*}

1 Université de Perpignan Via Domitia, Perpignan, France, **2** CNRS, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE), Perpignan, France, **3** Oregon State University, Corvallis, Oregon, United States of America, **4** CNRS, GenomiX IBISA, Montpellier, France, **5** Leibniz Institute for Age Research—Fritz Lipmann Institute, Jena, Germany

☞ These authors contributed equally to this work.

* christoph.grunau@univ-perp.fr



OPEN ACCESS

Citation: Roquis D, Lepesant MJ, Picard MAL, Freitag M, Parrinello H, Groth M, et al. (2015) The Epigenome of *Schistosoma mansoni* Provides Insight about How Cercariae Poise Transcription until Infection. PLoS Negl Trop Dis 9(8): e0003853. doi:10.1371/journal.pntd.0003853

Editor: Sergio Costa Oliveira, Federal University of Minas Gerais, BRAZIL

Received: November 26, 2015

Accepted: May 27, 2015

Published: August 25, 2015

Copyright: © 2015 Roquis et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: ChIP-Seq and RNA-Seq fastq files are available at the NCBI SRA as studies SRP034587 and SRP035609, respectively. We also provide a Gbrowse interface at <http://genome.univ-perp.fr>

Funding: The study was funded by the Agence Nationale de la Recherche, contract ANR 2010 BLAN 1720 01 (EPIGEVOL) ([http://www.agence-nationale-recherche.fr/programmes-de-recherche/recherches-exploratoires-et-emergentes/blanc-generalite-et-contacts/blanc-presentation-synthese-du-projet/?tx_lwmsuivibilan_pi2\[CODE\]=ANR-10-BLAN-1720](http://www.agence-nationale-recherche.fr/programmes-de-recherche/recherches-exploratoires-et-emergentes/blanc-generalite-et-contacts/blanc-presentation-synthese-du-projet/?tx_lwmsuivibilan_pi2[CODE]=ANR-10-BLAN-1720)).

Abstract

Background

Chromatin structure can control gene expression and can define specific transcription states. For example, bivalent methylation of histone H3K4 and H3K27 is linked to poised transcription in vertebrate embryonic stem cells (ESC). It allows them to rapidly engage specific developmental pathways. We reasoned that non-vertebrate metazoans that encounter a similar developmental constraint (*i.e.* to quickly start development into a new phenotype) might use a similar system. Schistosomes are parasitic platyhelminthes that are characterized by passage through two hosts: a mollusk as intermediate host and humans or rodents as definitive host. During its development, the parasite undergoes drastic changes, most notable immediately after infection of the definitive host, *i.e.* during the transition from the free-swimming cercariae into adult worms.

Methodology/Principal Findings

We used Chromatin Immunoprecipitation followed by massive parallel sequencing (ChIP-Seq) to analyze genome-wide chromatin structure of *S. mansoni* on the level of histone modifications (H3K4me3, H3K27me3, H3K9me3, and H3K9ac) in cercariae, schistosomula and adults (available at <http://genome.univ-perp.fr>). We saw striking differences in chromatin structure between the developmental stages, but most importantly we found that cercariae possess a specific combination of marks at the transcription start sites (TSS) that has similarities to a structure found in ESC. We demonstrate that in cercariae no transcription occurs, and we provide evidences that cercariae do not possess large numbers of canonical stem cells.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Conclusions/Significance

We describe here a broad view on the epigenome of a metazoan parasite. Most notably, we find bivalent histone H3 methylation in cercariae. Methylation of H3K27 is removed during transformation into schistosomula (and stays absent in adults) and transcription is activated. In addition, shifts of H3K9 methylation and acetylation occur towards upstream and downstream of the transcriptional start site (TSS). We conclude that specific H3 modifications are a phylogenetically older and probably more general mechanism, *i.e.* not restricted to stem cells, to poise transcription. Since adult couples must form to cause the disease symptoms, changes in histone modifications appear to be crucial for pathogenesis and represent therefore a therapeutic target.

Author Summary

The blood fluke *Schistosoma mansoni* causes intestinal bilharzia. The parasite has a complex life cycle in which a freshwater snail serves as intermediate host from which the human infecting larvae hatch. These larvae will actively seek skin contact, penetrate through the epithelium and start developing straight away into adult worms. Development from larvae into adults needs thorough adjustment of gene expression through repositioning or modification of proteins that are associated with DNA (the chromatin). We decided to compare the chromatin of human infective larvae (cercariae), the first developmental stage after infection of the vertebrate host (schistosomula) and adults of *S. mansoni*. We found that cercariae possess chromatin structures (modifications of histone H3) around the beginning of genes that are very different from schistosomula and adults. We conclude that this structure serves to keep gene transcription in a poised state, *i.e.* transcription is initiated and can start immediately when the blocking histone modification is removed. A similar type of histone modification was found in embryonic stem cells of vertebrates and our data indicate that it is either a more ancient and/or more general means to poise transcription than previously assumed. Since many parasites possess infective stages that develop rapidly within the host, this particular chromatin structure could be a therapeutic target for a new class of antiparasitic drugs.

Introduction

Histones are closely associated with DNA. Amino acids in their side-chains carry several modifications, such as methylation, acetylation and others. Specific combinations of histone modifications are associated with transcriptionally permissive or repressive chromatin structures, thus controlling gene expression at the transcriptional level [1]. Acetylation of histone H3 at lysine 9 (H3K9ac) is a mark associated with euchromatin and facilitation of transcription [2–6]. In model species the main enrichment of this mark is observed between the beginning of promoters and the 5' end of genes and sometimes in enhancers [2,3,7]. In contrast, methylation at the same position (H3K9me3) is a landmark of constitutive heterochromatin. It is often highly enriched at telomeric, pericentromeric and repeat-rich regions, and believed to be responsible for stable large-scale repression [8–11]. In *Plasmodium falciparum* for instance, it is found in subtelomeric and some small intrachromosomal regions containing almost exclusively clusters of genes involved in antigenic variations [6]. In human, it was shown that this

modification is mutually exclusive with trimethylation of histone H3 at lysine 4 (H3K4me3) on a given nucleosome [12]. H3K4me3 is often associated with transcription start sites (TSS), while trimethylation of H3 at lysine 27 is generally found in transcriptionally repressive heterochromatin. Consequently, the discovery of methylation at both sites (bivalent methylation) in several hundred genes of embryonic stem cells has attracted wide attention [13]. Embryonic stem cells (ESCs) are generated from the inner cell mass of mammalian blastocysts. These cells are self-renewing and can give rise to all lineages of the developing organism. The current view of bivalent methylation is that histone H3 trimethylated at lysine 27 (H3K27me3) represses transcription of lineage control genes during pluripotency, while H3K4me3 maintains them poised, *i.e.* transcription machinery is in position at the TSS, ready for activation upon reception of a signal that triggers differentiation (reviewed in [14]). When ESCs initiate differentiation, genes marked with bivalent trimethylation have two possible outcomes: H3K27me3 is removed and transcription occurs, or H3K4me3 is removed and genes stay silent. Since histones and histone modifications are extremely conserved through all kingdoms, we wondered if H3K9 methylation/acetylation and bivalent H3K4/K27 trimethylation might be a general mechanism by which a transient developmental state is maintained until development and differentiation into the next developmental state is triggered. Parasites are one of the models of choice to study this question since many of them have life cycles with very strong phenotypic differences between the stages, they are phylogenetically only distantly related to vertebrates and the environmental triggers are well-known, even though the molecular bases of switching to a new developmental stage are mostly unknown. Here we used the platyhelminth trematode *Schistosoma mansoni*. *S. mansoni* is the causative agent of intestinal bilharzia, a disease affecting 67 million people [15]. In contrast to many other parasites, the stage of the parasite infecting humans is actively seeking skin contact and penetrates the epidermis by use of cytolytic enzymes secreted from the pre- and post-acetabular glands and mechanical force. During skin penetration, cercariae lose their tails and undergo a drastic morphological and physiological transformation. Within two hours, the free-living larvae become obligatory endoparasitic schistosomula [16] that develop into adult worms, which reproduce in the blood stream of the host. Cercariae have therefore a double function: they are a vehicle for the parasites heritable information, but also true free-living animals with highly organized structures such as a nervous system and sensory organs, a digestive tract, excretory bladder and ducts, and a gland cell network. In addition to this, they have the capacity to develop very rapidly into full-grown adult worms, which are phenotypically very distinct from the cercariae.

The core histone isoforms of *S. mansoni* are canonical with 100% identity to human H4.A, 79–99% identity to H3 forms and 62–92% to H2A and H2B [17]. The distribution of histone modifications in *S. mansoni* was completely unexplored, and we decided to study the parasite epigenome by chromatin immunoprecipitation followed by sequencing (ChIP-Seq) of DNA associated with the above-mentioned histone isoforms H3K4me3, H3K9ac, H3K9me3 and H3K27me3. The genome of *S. mansoni* has been recently assembled and most of the gene structure data comes from automatic annotation [18]. Since gene prediction algorithms often miss the first exon (especially if they are untranslated), we performed RNA-Seq to unambiguously identify transcriptional start sites (TSS) and the expression status of genes.

Methods

Ethics statement

Housing, feeding and animal care followed the national ethical standards established in the writ of February 1st, 2013 (NOR: AGRG1238753A) setting the conditions for approval, planning and operation of establishments, breeders and suppliers of animals used for scientific

purposes and controls. The French Ministère de l'Agriculture et de la Pêche and French Ministère de l'Éducation Nationale de la Recherche et de la Technologie provided permit A 66040 to our laboratory for experiments on animals and certificate for animal experimentation (authorization 007083, decree 87–848) for the experimenters.

Animal breeding and preparation of parasites

The *S. mansoni* strain *Smbre*, originally sampled in Brazil was used in this study. The strain was maintained in its sympatric intermediate host strain *BgBRE* of the mollusk *Biomphalaria glabrata*, and *Mus musculus* as definitive vertebrate host. Cercariae were collected from infected snails 30 days after infection by pipetting from spring water and sedimentation on ice. For isolation of adults, infected mice were killed after 60 days of infection by a lethal intraperitoneal injection of sodium pentobarbital, and adult worms were recovered by retrograde perfusions of the hepatic portal system with citrate (7.5%) saline (8.5%) solution administrated through the left ventricle [19]. Worms trapped in the liver or mesenteric system were collected after excising these organs.

In-vitro transformation of cercariae to schistosomula was done as described before [20]. The shells of infected snails were cleaned to remove rotifers, transferred to de-chlorinated water (about 75 mL for 10 snails), and exposed to a light source (60 W) for two hours in order to promote cercariae emission. Cercariae were gently collected with a 1 mL micro-pipet previously moistened, avoiding snail feces and mucus. They were transferred to a sterile 50 mL conical tube and placed on ice in the dark. After 30 minutes, cercariae were concentrated by centrifugation (400×g/10 minutes/4°C) and the supernatant was removed, leaving 5 mL of liquid. Five milliliters of phosphate buffered saline, supplemented with 200 U/mL of penicillin and 200 µg/mL of streptomycin (PBS-PS) were added and put at 37°C. The tails of cercariae were sheared by vortexing at maximum speed during one minute and subsequently centrifuged (400×g/1 minute/4°C). All further handling was done under sterile conditions. The supernatant, containing only tails, was completely removed and discarded. The pellet was resuspended in 2 mL PBS-PS. Schistosomula were separated from tails by centrifugation along a gradient of polyvinylpyrrolidone-coated colloidal silica particles (Percoll) [21] prepared in 15 mL conical tubes (6 mL of the Percoll [Sigma, P4937], 1 mL 10 fold PBS, 0.25 mL 100 mM HEPES, 0.25 mL penicillin/streptomycin, 2.5 mL sterile water). One milliliter of the schistosomula suspension was applied to each column without disturbing the surface of the Percoll gradient. The preparation was centrifuged at 450×g for 10 minutes, with weakest possible brake. The top layer containing the tails was removed, and the pellet containing schistosomula was resuspended and washed twice in 10 mL of Dulbecco's modified Eagle's medium (DMEM) with 200 U/mL of penicillin, 200 µg/mL of streptomycin and 10 mM HEPES at 37°C (centrifugation 400×g/1 minute/4°C). Finally, schistosomula were resuspended in modified Basch's medium without antimycotics [20,22] and maintained at 37°C under 5% CO₂ for 21 hours.

RNA extraction and reverse transcription, quantitative PCR, and native chromatin immunoprecipitation

As the genome annotation of *S. mansoni* has mostly been done through prediction software, we performed RNA-Seq on cercariae to unambiguously identify TSS and exon-intron structures. Since published RNA extraction protocols for cercariae did not yield RNA suitable for Illumina sequencing, we developed the following method: 5,000 cercariae were ground with glass beads in liquid nitrogen for 5 minutes, resuspended in 0.5 mL TRIzol (Invitrogen), and 0.1 mL of chloroform was added. The tube was shaken vigorously for 1 minute and left at room temperature for 3 minutes, and was then centrifuged at ≥12,000×g for 15 minutes at 4°C. The

upper phase was transferred to a new RNase-free tube, an equal volume of 70% ethanol was added and the tube was vortexed. PureLink RNA Mini kit (Ambion, 12183020) was used for further purification following the manufacturer's protocol. RNA was recovered in 30 μ L RNAsure (Ambion, AM7006), left for 1 minute at room temperature, collected by centrifugation for 1 minutes at 8,000xg at room temperature and incubated at 65°C for 10 minutes. To remove remaining DNA, 3 μ L of 10X TURBO DNase Buffer (TURBO DNA-free, Ambion, M1907) were added to 30 μ L of total RNA in RNAsure, followed by 3 μ L of TURBO DNase (2 Units/ μ L). The preparation was gently mixed and incubated at 37°C for 30 minutes. DNase was removed with an RNA Clean-up system (RNeasy mini kit, QIAGEN, 74104) and eluted in 30 μ L RNase-free water. For schistosomula, total RNA extraction, DNase treatment and purification methods were adapted from the protocol described above. 500 ng of this purified total RNA were reverse transcribed using identical concentrations (250 nmol/L) of random hexamers and oligo(dT)₁₈ primers of Maxima H Minus Reverse Transcriptase kit (Thermo-SCIENTIFIC). Quantitative PCR was performed using a LightCycler 480 System (Roche Diagnostics). We used 2.5 μ L of cDNA (diluted 1/10 in water) in a total volume of 10 μ L containing 1X LightCycler 480 SYBR Green I Master Mix (Roche Diagnostics) and 500 nmol/L of each primer. To ensure that a single product was generated, we separated amplification products on a Labchip GX DNA assay system (PerkinElmer). For each studied stage, the level of mRNA was normalized using the mean geometric transcription rate of three reference sequences (Smp_093230, Smp_197220 and Smp_089880) previously described by Wang and colleagues [23]. mRNA of Smp_034000 was measured using primers Smp034000.2F: GCGTGCTGTGG AGGAAAGCGA and Smp034000.2R: ACGCAGACAGGGCATCGTCA. The normalized relative quantities (NRQs) were calculated as described by Hellemans and colleagues [24], using equation 1 where E_{target} is the amplification efficiency of the gene of interest; E_{ref} is the amplification efficiency of the reference gene; $\Delta C_{p,\text{ref}} = C_{p,\text{ref}}(\text{calibrator}) - C_{p,\text{ref}}(\text{sample})$; and $\Delta C_{p,\text{target}} = C_{p,\text{target}}(\text{calibrator}) - C_{p,\text{target}}(\text{sample})$. For each reaction, the crossing point (Cp) was determined using the second derivative maximum method applied using Light Cyclers Software version 3.3 (Roche Diagnostics). Male cercariae were arbitrarily chosen as calibrator. We obtained male cercariae by infecting host mollusk *B. glabrata* strain BgBre with single miracidium from SmBre strain. Sex determination was performed as described in [25]. Experiments were done in six biological replicates.

Native chromatin immunoprecipitation was done according to an earlier published protocol [26] and at least 1,500 cercariae, 2000 schistosomula or 1 adult couple using the following antibodies: H3K4me3 (Millipore, cat# 04-745 lot# NG1680351, 4 μ L per reaction), H3K9Ac (Millipore, cat# 07-352 lot# DAM16924924, 8 μ L per reaction), H3K9me3 (Abcam, cat# ab8898 lot# 733951, 4 μ L per reaction), and H3K27me3 (Diagenode, cat# pAb-069-050 lot# A29900242, and cat# C15410069 lot# A1821D, 8 μ L per reaction). For each sample, we used a control without antibody, to assess unspecific background (bound fraction) and input (unbound fraction). Inputs were used for normalization in all subsequent bioinformatics analyses. Experiments were done in duplicate. Antibodies were carefully tested for specificity as described in [27] and saturating quantities (see [27]) were used. Further details are available at <http://methdb.univ-perp.fr/epievo/>

Illumina library generation and sequencing

For ChIPseq library construction, the ChIP sample preparation kit (Illumina Inc., USA) was used according to the manufacturer's protocol except for the indexed adapters. Those adapters came from the Truseq DNA sample preparation kit (Illumina Inc., USA). Briefly, 30 ng of DNA per condition was blunt ended and adenylated on its 3' ends. Illumina's indexed adapters

were ligated to both ends of the adenylated DNA. Ligated DNA was size separated by electrophoresis through a 2% agarose gel and a size selection was performed at 400 bp. Once extracted from the agarose, size selected DNA was enriched by 18 cycles of PCR for 10 seconds at 98°C, 30 seconds at 60°C and 30 seconds at 72°C using Illumina's proprietary primers and Phusion DNA polymerase (New England Biolabs Inc., USA). Each library was verified using a DNA 1000 Chip on a Bioanalyzer 2100, quantified by real time PCR with the KAPA Library Quantification Kit for Illumina Sequencing Platforms (Kapa Biosystems Ltd, SA) and then diluted to 10 nM.

RNAseq library construction was done with the TruSeq RNA Sample Seq kit (Illumina Inc., USA) according to the manufacturer's protocol. Briefly, poly-A containing mRNA molecules were purified from 70 to 450 ng of total RNA using oligo-dT magnetic beads. The purified mRNA was fragmented by addition of the fragmentation buffer and was heated at 94°C in a thermocycler for 8 minutes. First strand cDNA was synthesized using random primers. Second strand cDNA synthesis, end repair, A-tailing, and adapter ligation was done in accordance with the manufacturer's instructions. Purified cDNA templates were enriched by 15 cycles of PCR for 10 seconds at 98°C, 30 seconds at 65°C, and 30 seconds at 72°C using Illumina's proprietary primers and Phusion DNA polymerase. Each indexed cDNA library was verified using a DNA 1000 Chip on a Bioanalyzer 2100 and quantified by real time PCR with the KAPA Library Quantification Kit for Illumina Sequencing Platforms (Kapa Biosystems Ltd, SA) then diluted to 10 nM using Illumina's resuspension buffer.

For each lane of sequencing, four libraries (ChIPseq) or 6 libraries (RNAseq) were pooled in equimolar concentrations, denatured using 0.1N NaOH and diluted to a final concentration of 7 pM (ChIPseq) or 8 pM (RNAseq) using Illumina's HT1 buffer. 120 µL of the dilution was then transferred into a 200 µL strip tube and placed on ice before loading onto the cBot. The flow cell was clustered using Single Read Cluster Generation Kit (Illumina Inc., USA), according to the Illumina SR_amplification_Linearization_Blocking_PrimerHyb recipe. The flow cell was loaded onto the Illumina HiSeq 2000 instrument following the manufacturer's instructions and sequencing was performed with the 50 cycles, single read, indexed protocol. Image analyses and base calling were performed using the HiSeq Control Software (HCS) and Real-Time Analysis component (RTA). Demultiplexing was performed using CASAVA (Illumina). The quality of the data was assessed using fastqc from the Babraham Institute and the Illumina software SAV (Sequence Analysis Viewer).

Illumina sequencing quality control, alignment and ChIP-Seq peak calling

All data treatment was carried out under a local galaxy instance [28] (<http://galaxy.univ-perp.fr>). Fastq Groomer was used for verification of the fastqsanger format, and the FASTX-Toolkit (Compute quality statistics, Draw quality score boxplot, Draw nucleotides distribution chart) was used for initial quality control. Read quality was judged sufficiently good (the majority of reads showed a quality score ≥ 24 for all positions) and no further quality filter was applied.

For ChIP-Seq, reads were aligned to the *S. mansoni* reference genome, assembly version 5.0 (downloaded from <ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/Assembly-v5/>) using Bowtie2 [29] evoking parameters—sensitive-k 2. Mapping quality (mapq) in Bowtie2 is related to “uniqueness” of the read. The higher the quality score is, the lower is the probability that a read can map elsewhere in the genome. By filtering mapped reads with a score of 255 (the highest possible) in samtools [30] (samtools view-hs-q 255), we kept reads with a single match on the genome. In Bowtie 2.1 score 255 was abandoned and filtering of uniquely mapped reads was done with “samtools view-Sh-q *quality value* 40–42-F 0x0004 - |

grep -v XS:i. ". SAM alignment files were converted into the bed format with pyicos [31] and sorted with sortBed-i of the bedtools suite [32]. For peak identification an equal number of random lines in the bed-file was chosen for each biological replicate (Table 1). Although comparative analysis software usually include a normalization step to handle the different numbers of mapped reads in conditions or replicates when comparing samples, we found that the random sampling step drastically reduces the amount of false positives without lowering the sensitivity. Mostly, this removes background peaks. Identification of peaks was done with ranger of Peak-ranger v1.16 [33] with P value cut off 0.0001, FDR cut off 0.01, Read extension length 200, Smoothing bandwidth 99 and Delta 1. We used the input samples as negative controls for the peakcalling (-c). For data visualization, wiggle files were generated and uploaded to a local GBrowse instance [34] (<http://genome.univ-perp.fr>).

ChIP-Seq results in adults and cercariae were verified by qPCR (primers in S1 Table) in 20 randomly chosen regions using the following conditions: denaturation 95°C for 10 seconds, annealing 60°C for 10 seconds, extension 72°C for 20 seconds, over 40 cycles. The kit used for the amplification is the LightCycler 480 SYBR Green 1 Master (Roche, 04887352001) and the instrument is a LightCycler 480 (Roche). To compare ChIP-Seq and qPCR results, coverage was calculated with coverageBed of the bedtools suite for the regions whose boundaries were defined by the corresponding qPCR primer positions and divided by the coverage in the alpha-tubulin reference locus (Smp_scaff000423:21,107–21,289). Spearman rank correlation gave values > 0.8, R2 was 0.64 for H3K4me3 and 0.59 for H3K27me3 indicating that both methods gave comparable results (S1 Fig).

For RNA-Seq, reads were aligned to the genome with TopHat v1.4.1 with default parameters and exon-intron structure was reconstructed with cufflinks v1.3.0 evoking-q-I 40000-F 0.100000-j 0.150000-p 8 on each library and data were merged and linked to the gene prediction of the Sanger Centre (<ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/Assembly-v5/>) with Cuffmerge v1.0.0 [35].

Comparative ChIP-Seq analysis

To identify peaks that were different between cercariae and adults, we used the DiffBind library 1.6.1 [36] available on Bioconductor [37] for R version 2.15.1. As input, we used the randomly sampled aligned reads that also served as inputs for PeakRanger and the regions file generated with PeakRanger, all in bed format. Diffbind allows for replicates and "input" negative control to be taken into account to search for differential peak distribution between conditions. For H3K9ac, PCR duplicates were removed with pyicos, which improved sample autocorrelation. For H3K4me3 all differences were inspected on GBrowse. For the other histone isoforms, one hundred regions identified by DiffBind were randomly sampled with a custom Perl script and verified by visual inspection. Average histone modification profiles around TSS were generated by doing a "window analysis" around the TSS of genes, using EpiChIP v0.9.7-e [38]. EpiChIP allows for generating genome-wide, average profiles for histone modifications at a given genomic feature (in our case, around the TSS for genes). For this purpose, we used the output file of Cuffmerge and selected only those 12,871 genes located on linkage groups that were assembled at the chromosome level. Chromosome names were converted in the input bed files into EpiChIP canonical names ("Chrnumber") with a custom script. EpiChIP was also used to correlate transcription and histone modification, and to quantify the proportion of peaks that contribute significantly to the average profiles. Common peaks were detected from Peakranger region bed-files with multiIntersectBed of the BedTools package.

For the analysis of chromatin structure around repetitive sequences, reads were aligned to a fasta file containing all 3,145 known repeats of *S. mansoni* as consensus sequences (available

Table 1. Summary of next-generation sequencing data and analysis results.

Condition	H3K4Me3			H3K9Ac		H3K9Me3		H3K27Me3		
	cercariae	adults	schistosomula	cercariae	adults	cercariae	adults	cercariae	adults	schistosomula
Number of samples	2	2	2	2	2	2	2	3	3	2
Input reads (average per sample)	34,519,715	6,421,317	49,109,097	54,833,141	43,884,060	53,702,142	43,884,060	40,895,894	33,865,527	46,158,137
Alignment to unique sequences										
Aligned exactly 1 time (average per sample)	21,455,086	3,097,077	28,275,831	28,626,785	28,035,327	23,326,061	21,945,541	22,882,234	27,123,469	30,651,658
Aligned reads used for peak-calling (per sample)	2,900,000	2,900,000	2,900,000	23,000,000	23,000,000	21,000,000	21,000,000	17,900,000	17,900,000	17,900,000
Regional peaks identified by peakranger 1.16 (average for condition)	7,746	8,868	7,376	17,116	421	128,518	100,884	145,322	27,544	65,629
Significantly different peaks identified by DiffBind	224			3,844		9,505		8,480		
DiffBind result confirmed by visual inspection	124			random sample of 100 peaks for each condition(33–50% confirmed by 2 independent observers)						
EpiChIP peaks with signal at FDR 0.01 (per condition)	3,598	3,566	3,954	4,8014,3705,024						
Alignment to repetitive sequences										
Aligned exactly 1 time (average per sample)	1,326,620	242,000		3,805,162	3,091,945	3,881,193	3,612,100	3,999,829	4,554,097	
Random sample used for DESeq	220,000	220,000		2,900,000	2,900,000	3,400,000	3,400,000	3,900,000	3,900,000	

doi:10.1371/journal.pntd.0003853.t001

<http://methdb.univ-perp.fr/downloads/>) [39] using Bowtie2 [29] evoking parameter–k 2 and parameter preset “sensitive”. Unique matches were filtered using samtools [30] (mapping quality mapq = 255). In order to obtain identical effective library sizes for each sample, the same number of aligned reads was randomly sampled for each of the ChIP experiments (Table 1). After the counts had been normalized, they were compared between samples using the DESeq

package of the R software [40]. This package uses the negative binomial distribution to calculate the P-value for each element between two compared samples and find elements that present significant differences in their read enrichment levels.

In-situ detection of replication and transcription

For the detection of mRNA synthesis, cercariae were incubated in spring water at 28°C with a concentration of 1mM of 5-ethynyl-uridine (EU, Invitrogen Click-iT RNA Imaging Kit, MP10329) for 2, 18 h and 21 h. Freshly transformed schistosomula were grown in Basch medium at the same EU concentration in an incubator at 37°C and 5% CO₂, also for 2, 18, and 21 h. Labeling of replicating cells was performed in the exact same conditions, only substituting EU by a concentration of 10 mM of 5-ethynyl-2'-deoxyuridine (EDU, Invitrogen Click-iT DNA Imaging Kit, MP10337). Negative controls without the addition of 5-ethynyl-uridine (EU) and 5-ethynyl-2'-deoxyuridine (EDU) were also performed for every condition. Fixation, development and Hoechst 33342 coloration after incubation were performed following the protocols of the supplier. Imaging was done on a confocal microscope Zeiss LSM 700 at 488 nm (EU and EDU) and 405 nm (Hoechst 33342). For every condition, 100 animals were observed for transcription and replication.

Accession numbers

RNA-Seq and ChIP-Seq reads are available at the NCBI-SRA under study accession numbers SRP034587 (ChIP-Seq) and SRP035609 (RNA-Seq).

Results

Chromatin landscape in *S. mansoni* cercariae and adults

To link transcription and chromatin structure, we considered the correct identification of the TSS as crucial. Anecdotal evidences had suggested that the available automatic annotation had missed a certain proportion of first exons and we decided to perform RNA-Seq with RNA extracted from cercariae. Indeed, our data showed that in about 20% of the genes, TSS had to be corrected. We then performed ChIP-Seq. For H3K4me3, we found relatively sharp peaks with a mean peak maximum located 250 bp (or 1–2 nucleosomes) downstream of the transcription start site (TSS) of genes (Fig 1A). We detected on average 7,746 peaks in cercariae and 8,868 peaks in adults (Table 1). For about ~24% of genes, EpiChIP identified transcripts but no significant H3K4me3 peaks (FDR = 0.01). Examples for genes that are free of H3K4me3 peaks are all known micro exon genes (MEGs) supplementary table S5.4 in [41] (Fig 2A), with the exception of Smp_163710 that shows small peaks in adults. MEGs contain unusually small exons of 3–36 bp that are repeated in tandem [42]. Their function remains elusive but it is thought that hypervariable proteins can be generated by MEGs through a 'pick and mix' splicing strategy. It is conceivable that such types of genes have a specific control mechanism. Other H3K4me3-peak free genes are polycistronic genes that are located downstream of the trans-splicing site. Protasio *et al.* estimated trans-splicing events to occur in 11% of *S. mansoni* genes [18] and had compiled a list of putative polycistronic transcripts. We used this list and show that in 143 (89.9%) genes, the H3K4me3 peak is located exclusively in the gene upstream of the trans-splicing acceptor site (Fig 2B), in 2 (1.2%) cases peaks are present in both genes (upstream and downstream), and in 13 (8.2%) genes we did not find any peak, although in 3 cases we suspect that an upstream gene with a peak may be the first of the polycistronic unit (S2 Table). Our data are in line with the view that these genes are indeed transcribed as one unit from a single TSS (that has the H3K4me3 signal) and fragmented by trans-splicing. For

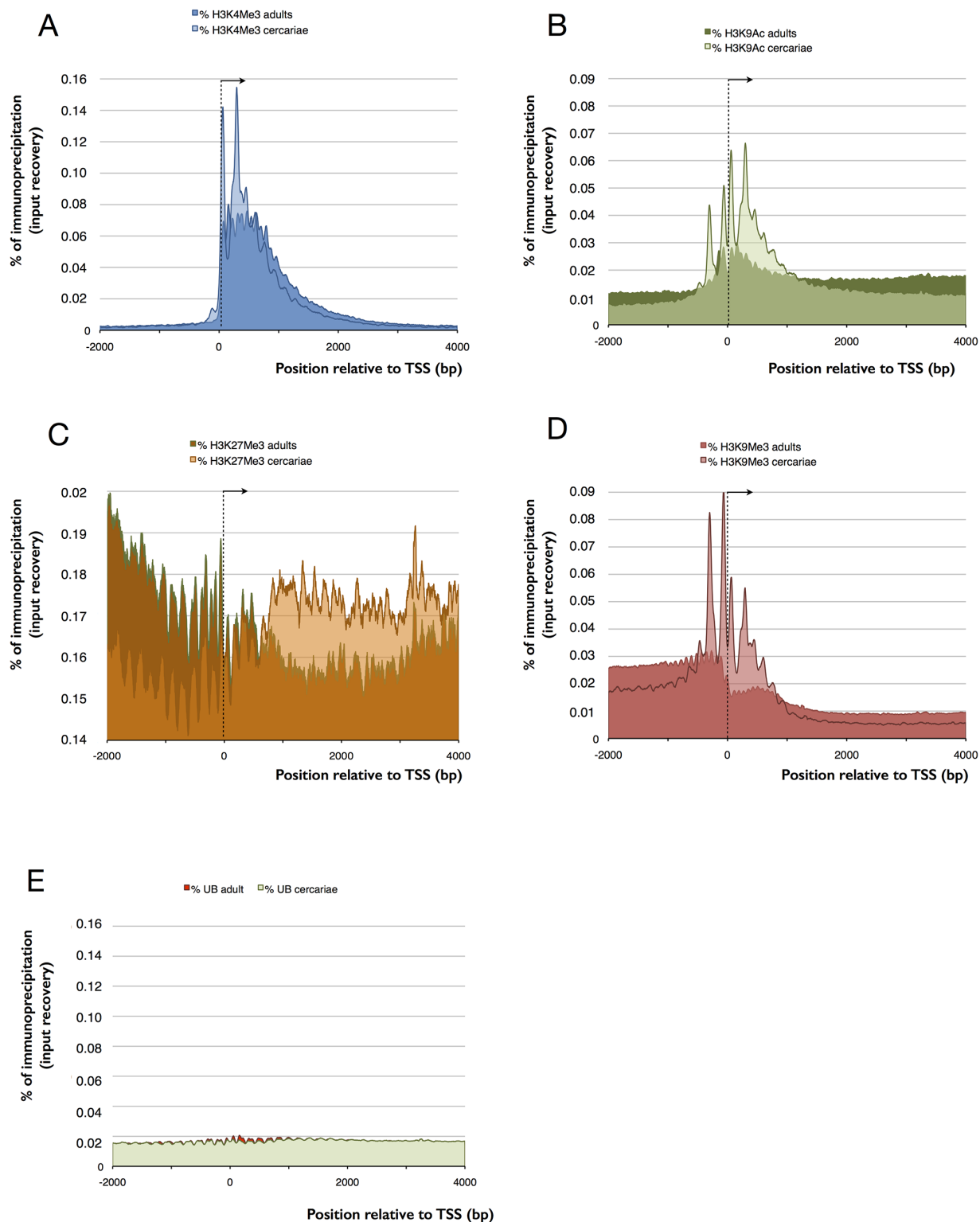


Fig 1. Comparison of EpiChIP profiles 2kb upstream and 4kb downstream of the TSS (average for 12,871 genes located on linkage groups that are assembled on chromosome level) for cercariae (light color) and adults (dark color). TSS and direction of transcription indicated by an arrowhead. X-

axis in bp, y-axis in relative enrichment (% of all mapped reads). (A) H3K9me3, (B) H3K9ac, (C) H3K27me3, (D) H3K9me3, and (E) unbound fractions UB of mock ChIP without antibody that corresponds to input that shows no difference between cercariae and adults and no enrichment at the TSS.

doi:10.1371/journal.pntd.0003853.g001

H3K27me3 we found with Diagenode antibody cat# pAb-069-050 lot# A29900242 enrichment at the TSS (-500 to +1,000 bp), while it was reduced to background level in adults. Since there was anecdotal evidence for cross-reactivity of this antibody with H3K4me3, we repeated the experiment with cat# C15410069 (that replaces pAb-069-050) lot# A1821D. With this antibody we also observe enrichment of H3K27me3 in cercariae compared to adults, but shifted towards the body of the gene (Fig 1C and Fig 2C). Methylation and acetylation in H3K9 are

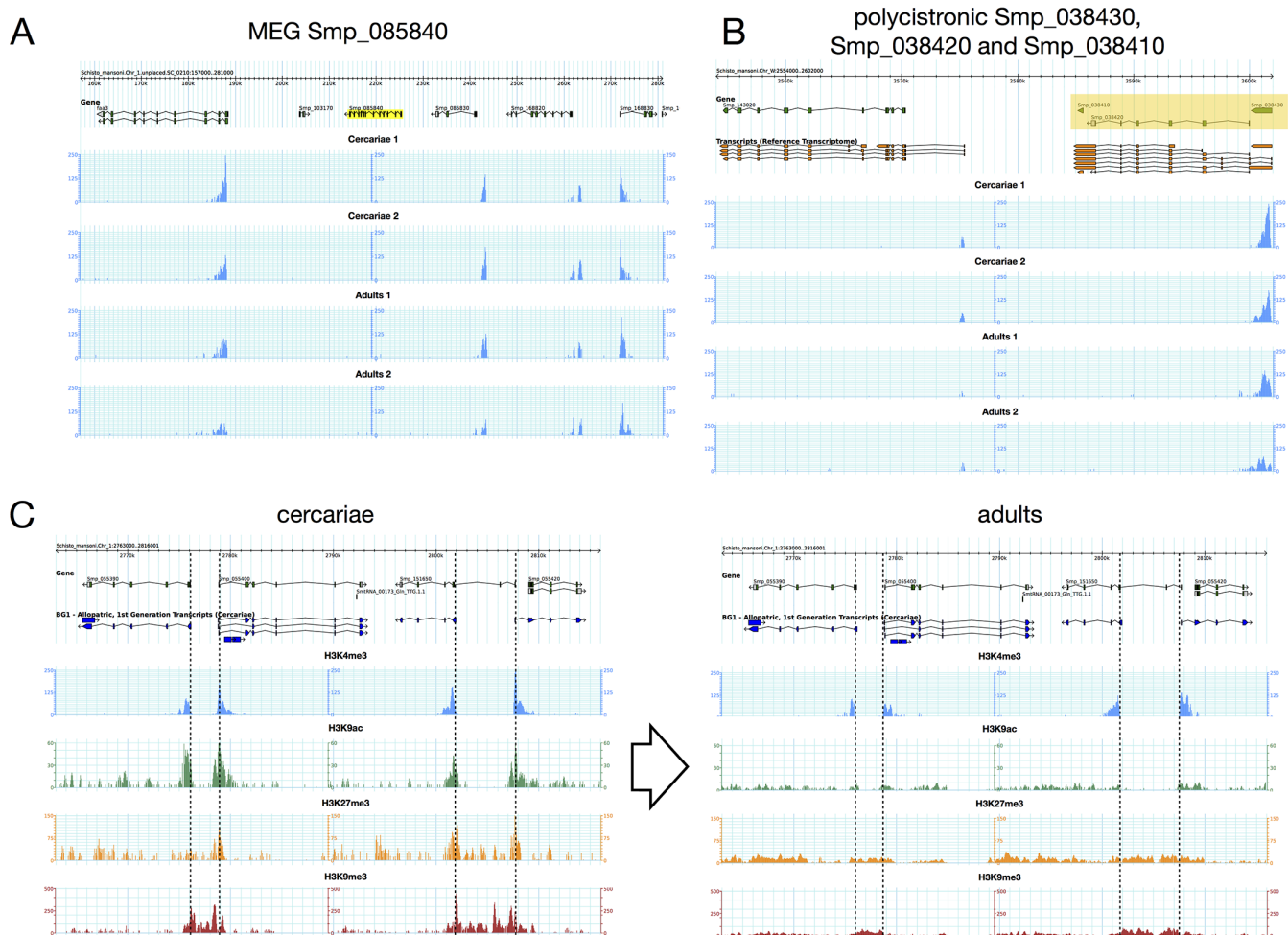


Fig 2. Typical examples of chromatin profiles. (A) H3K4me3 peaks are absent in the TSS of MEG Smp_085840 (highlighted in yellow) while surrounding genes show characteristic peaks in all 4 samples. (B) Polycistronic genes Smp_038430, Smp_038420 and Smp_038410 (from right to left, highlighted in yellow) show an H3K4me3 peak only in the first gene. (C) Changes in chromatin structure in cercariae (left) and adults (right). On top predicted genes, below RNA-Seq based exon/intron structure, followed by H3K4Me3 (blue), H3K9Ac (green), H3K27Me3 (yellow) and H3K9Me3 (red), identical scale for cercariae and adults. Positions of TSS that were detected by RNA-Seq are indicated by a dotted line.

doi:10.1371/journal.pntd.0003853.g002

mutually exclusive at the same histone and H3K9me3 is considered a repressive mark. H3K9me3 shows a peak about 200 bp upstream of the TSS in cercariae, which is reduced to an upstream background in adults (Fig 1D and Fig 2C). H3K9ac is a hallmark of active transcription. In cercariae we detected this modification primarily around the TSS (-500 to +1,000 bp) while in adults acetylation spans the body of the genes (Fig 1B and Fig 2C). Taken together, our results indicate that all four studied histone modifications show an unusual enrichment around the TSS in cercariae, while in adults their distribution is similar to that found in other metazoans. We saw no differences between the input fractions (negative controls used for normalization) of cercariae and adults (Fig 1E).

In cercariae and adults, both H3K4me3 and H3K9ac showed globally low association with repeats ($4.10 \pm 0.43\%$ and $6.45 \pm 0.65\%$, respectively of the total reads aligned) (details in Table 1). For H3K27me3 and H3K9me3 the percentage was higher with $7.60 \pm 1.00\%$ and $7.75 \pm 0.60\%$. DESeq [40] identified repeats that change chromatin structure during transformation from cercariae to adults. DESeq uses a negative binomial distribution to model read counts that can be assigned to a locus. To estimate variance within a sample it pools the data from genes with similar hit counts, thus allowing variance estimation and identification of significant differences based on a small number of biological replicates [40]. We used DESeq because we had earlier compared it to edgeR [43], and DESeq2 and in our hands it gave the most consistent results. Using a p-value of $< 10^{-5}$ as the selection criterion, we identified 17 (0.54%) repeats that change their chromatin status for H3K4me3, 182 (5.79%) for H3K9ac, 263 (8.36%) for H3K9me3, and 40 (1.27%) for H3K27me3. Results are summarized in S3 Table. Interestingly, not all repeats were concerned equally: in roughly two thirds of these repeats ($n = 363$, 11.5%) more than one histone isoform had changed meaning that about 10% of all repeat families undergo important chromatin structure changes during the transformation of cercariae into adults. Two results were particularly interesting (Fig 3). The first is a chromatin change in five regions that are composed of tandemly repeated stretches of class I LTR retrotransposons and class II DNA transposons. These regions are located on the large *S. mansoni* chromosomes (1 to 4), but there might be similar regions on the three smaller chromosomes that were overlooked. Each of the regions is composed of a different repeat, but they all belong to the same repeat class (*i.e.* transposon). In these five regions, enrichment of histone forms H3K9me3 and H3K27me3 decreases during cercariae to adult metamorphosis, while enrichment of H3K9ac and H3K4me3 increases. Considering five states (H3K4me3 up, H3K9ac up, H3K9me3 down, H3K27me3 down, transposon or other), the individual term binomial distribution probability to obtain such a combination by chance is 3.7^{-36} . Our data suggest therefore that the process is not randomly distributed along the genome. Stretches of (retro)transposons could form heterochromatic knobs in the cercariae and their change in chromatin structure could be an indication of large scale chromatin rearrangements or chromosome displacements in the nucleus during transformation into adults. Another peculiar observation during transformation from cercariae to adults is the nearly complete loss of H3 modifications in 22 of the 36 W-specific repeats (61%). We had shown earlier that transcription of these repeats can only be observed during larval stages and is undetectable in adults [44]. From qPCR data for a small subset we had already postulated that these repeats acquire an “adult-specific” chromatin structure. This view is confirmed by the new NGS data: H3K9/K4/K27 methylation and H3K9 acetylation are lost. Cytogenetic images suggest that the repeat containing large pericentromeric regions becomes heterochromatic when cercariae develop in the snail, but no clear data is available for adult worms [45]. It is tempting to speculate that H3 is replaced by another H3 isoform similar to what is observed in centromeres and neo-centromeres of other species for instance [46]. A homologue of the centromere specific form of H3, CENP-A, was detected in *S. mansoni*

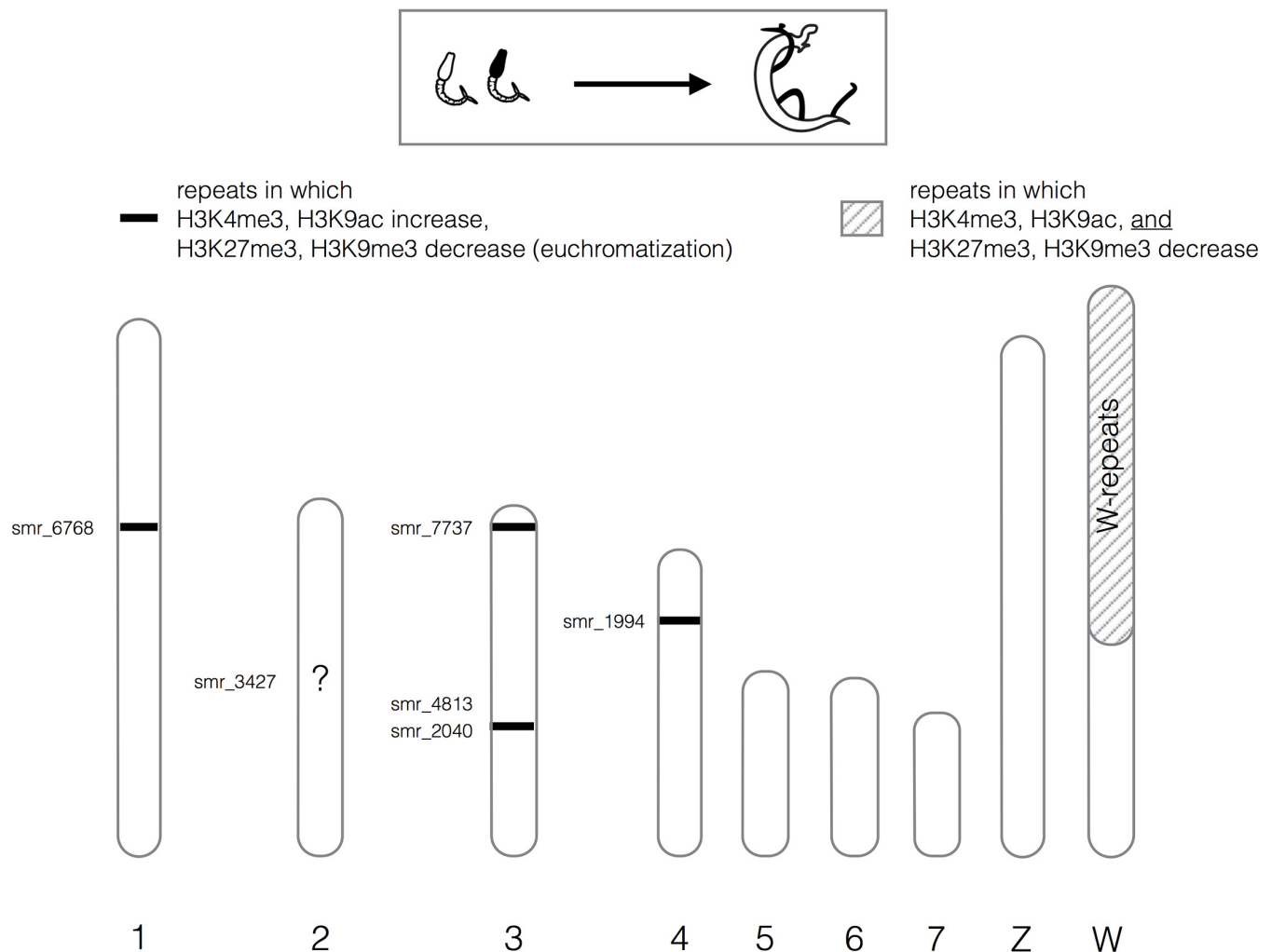


Fig 3. Schematic representation of chromatin structure changes in tandem-repeat regions during metamorphosis of cercariae into adults. Autosomes 1–7 are labeled with numbers, sex chromosomes with letters. Repeat regions are indicated on each chromosome as inferred by alignment to the reference genome, or previously by FISH on the W-chromosome. Repeat class smr_3427 is presumably on chromosome 2 but the exact position is unknown (unpositioned contig).

doi:10.1371/journal.pntd.0003853.g003

(Smp_040020) [47] but homology with CENP-A of other species is vague and basically restricted to the core H3/H4 domain.

Bivalent trimethylation of H3K4 and H3K27 is specific for a subset of genes in cercariae

While PeakRanger identify 1,122 more peaks in adults than cercariae for H3K4me3, Diffbind considers only 224 differences to be statistically significant. After visual inspection, we confirmed differential enrichment in only 124 regions along the genome with 95 (77%) hyper-methylated in cercariae and 29 (23%) in adults. This means that less than 2% of H3K4me3 peaks are different, while gene transcription patterns of cercariae and adults are clearly distinct [48]. A peculiar feature of the *S. mansoni* epigenome was revealed when we analyzed the

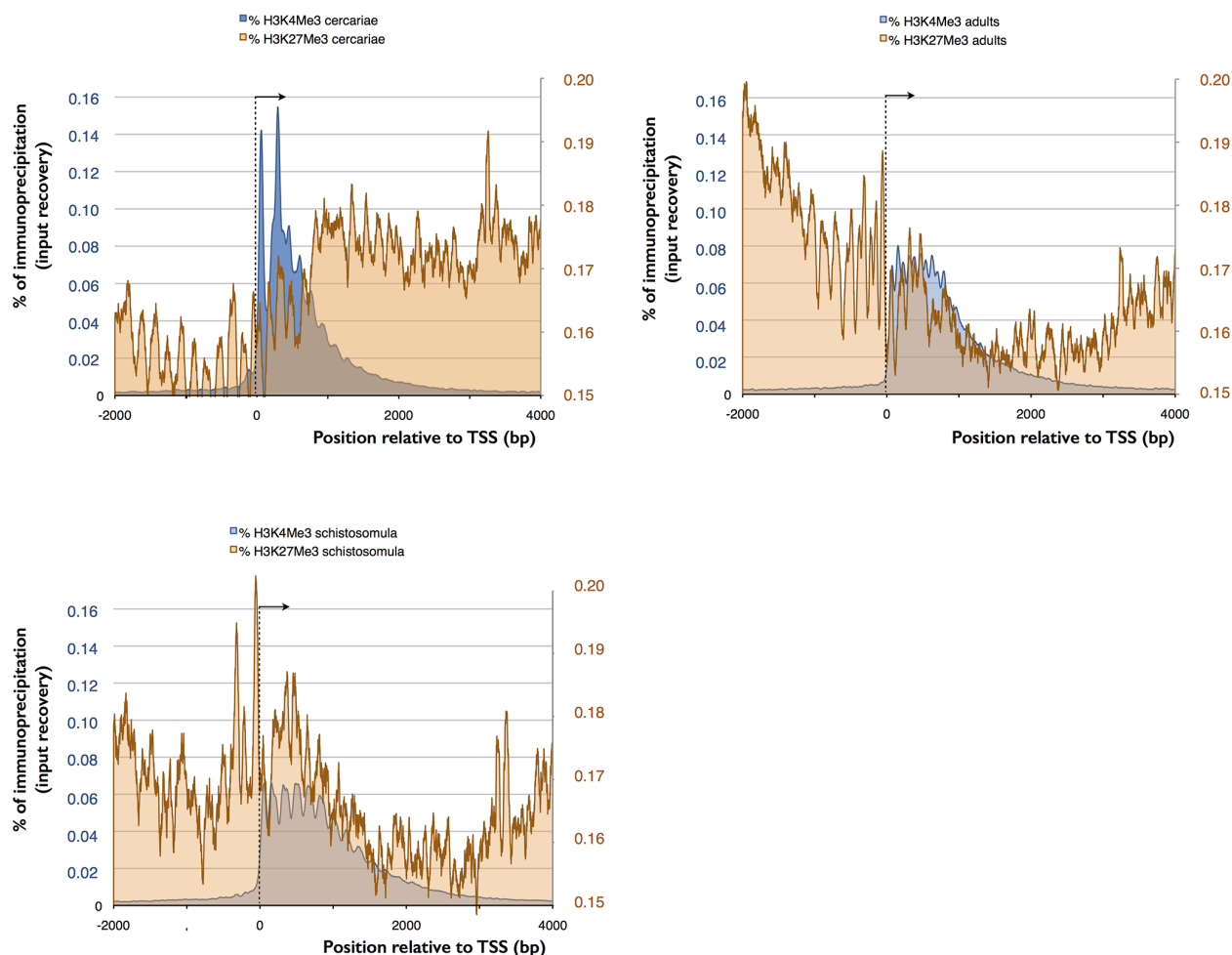


Fig 4. Comparison of EpiChIP profiles 2 kb upstream and 4 kb downstream of the TSS (average for 12,871 genes located on linkage groups that are assembled on chromosome level) between different histone isoforms in cercariae (left up), schistosomula (left down) and adults (right). TSS and direction of transcription indicated by an arrowhead. Bivalent methylation of H3K4 and H3K27 occurs in cercariae but not in adults, left y-axis for H3K4me3, right y-axis for H3K27me3.

doi:10.1371/journal.pntd.0003853.g004

distribution of the heterochromatic H3K27me3. In cercariae, H3K27 trimethylation is high in the gene body (Fig 4). H3K27me3 and H3K4me3 are also statistically significantly enriched around the TSS in 150 genes. This number was reduced to 121 after visual inspection (S4 Table). Thirty-five days of development separate cercariae and full-grown adults worms. To determine more precisely the time window during which chromatin is remodeled we performed ChIP-Seq on schistosomula. Our results show that the adult chromatin structure is already established in schistosomula by 18 h after infection (Fig 4). Most importantly, in schistosomula, bivalent methylation is absent in the 121 genes that were significantly enriched in cercariae.

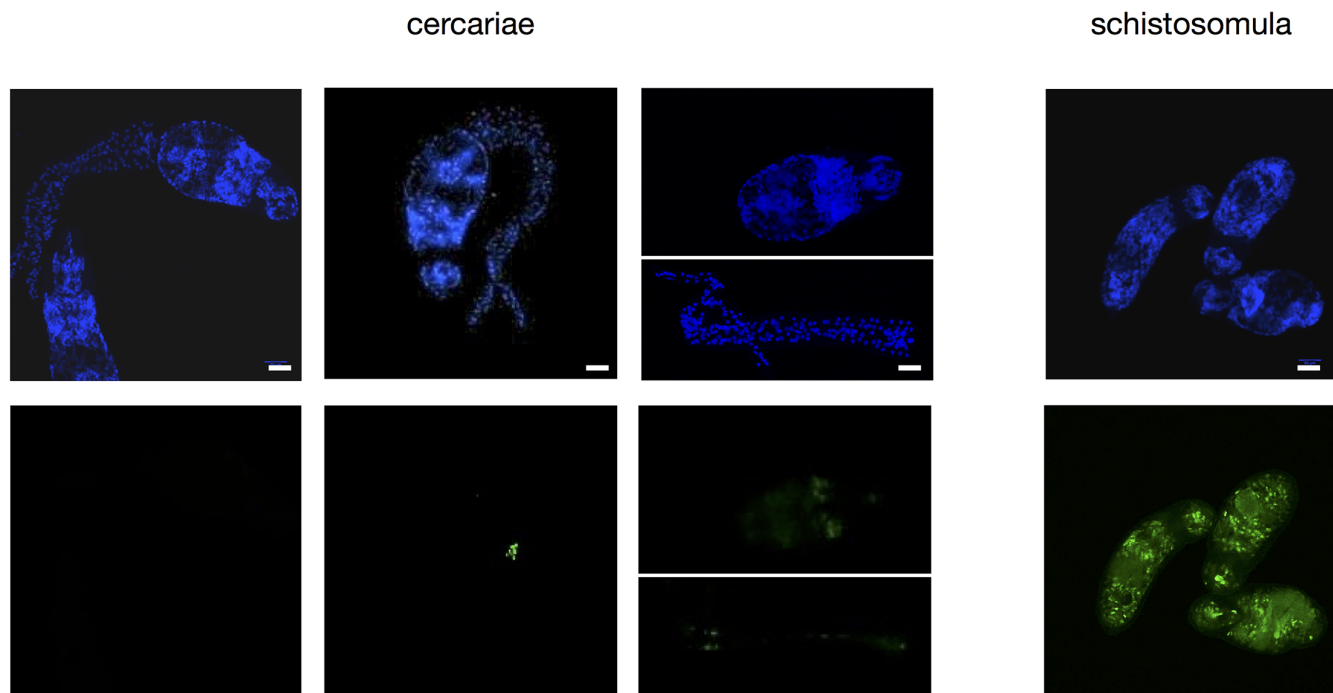


Fig 5. Detection of RNA synthesis by incorporation of EU by fluorescence microscopy. Top panel (blue) Hoechst 33342 staining of DNA, bottom panel (green) EU incorporation and detection with Alexa Fluor 488. White bar 20 μ m.

doi:10.1371/journal.pntd.0003853.g005

Cercariae are transcriptionally silent

While RNA is clearly present in cercariae, the co-localization of all four histone modifications at the TSS made us wonder if transcription is indeed ongoing or if it is poised. Cercariae are short lived and they do not grow outside the host. It is therefore conceivable that their RNA was produced during their development in the snail. We used incorporation of the cytosine analog 5-ethynyl-uridine (EU) and fluorescence labeling to detect active transcription *in vivo*. At 28°C, the maximum life span of a cercaria is 32 h (90% of mortality occurs past 26 h [49]) but infection capacity decreases rapidly and probably does not exceed 6 h. We decided therefore to use a relatively long EU pulse of 18 h. Despite this, we detected fluorescence signals in less than 10% of the cercariae. With the exception of a single individual (out of one hundred) signals were detected in only a few of the roughly 1,000 cells of an individual larva, and were always restricted to the nucleus (examples shown in Fig 5). In contrast, schistosomula, the first developmental stage after infection of the vertebrate host, showed consistent staining in the 100 arbitrarily chosen individuals and staining was more dispersed. We conclude that there is very little transcription in cercariae and that transcription is rapidly activated upon transformation into schistosomula. Our observation that in cercariae the repressive mark H3K9me3 is slightly upstream and the permissive mark H3K9ac is slightly downstream but close to the TSS is in accordance with stalled transcription. In contrast, in adults H3K9 acetylation and methylation are spread out with H3K9ac occurring along the gene body and H3K9me3 in the upstream regions separating genes (Fig 6A).

Bivalent H3K4/H3K27 trimethylation and absence of transcription are reminiscent of what occurs in ESC of vertebrates, e.g. mice [13], human [50] and zebrafish [51]. There, bivalent

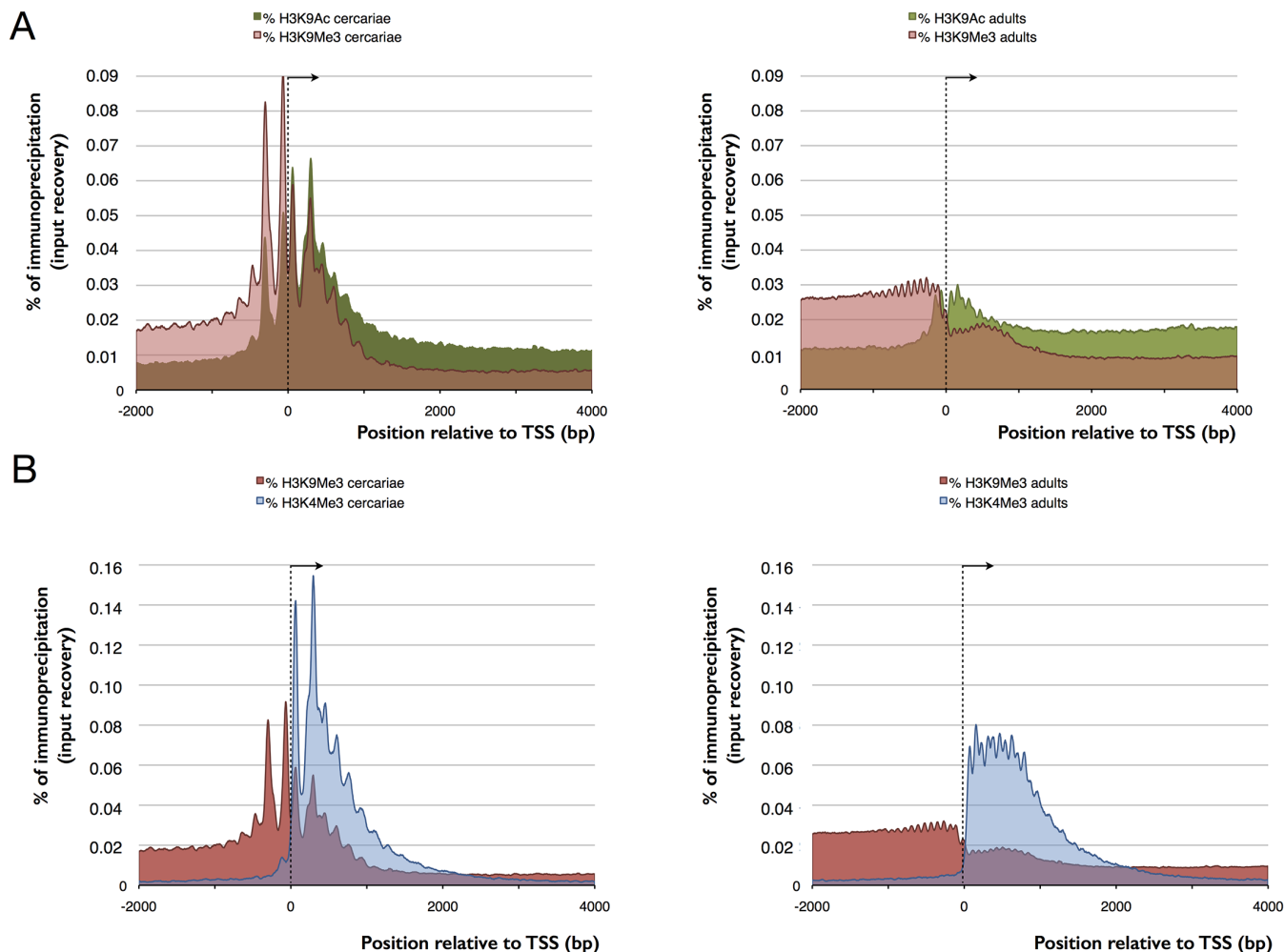


Fig 6. Comparison of EpiChIP profiles 2 kb upstream and 4 kb downstream of the TSS (average for 12,871 genes located on linkage groups that are assembled on chromosome level) between different histone isoforms in cercariae (left) and adults (right). TSS and direction of transcription indicated by an arrowhead. (A) methylation of H3K9 upstream of TSS and acetylation downstream, concentrated around TSS in cercariae and spanning the gene body (H3K9ac) and the intergenic region (H3K9me3) in adults, (B) methylation of H3K9 and H3K4 are mutually exclusive in adults but overlap in cercariae.

doi:10.1371/journal.pntd.0003853.g006

H3K4/H3K27 trimethylation occurs at the transcription start site of lineage-specific genes and it was proposed that it keeps them transcriptionally “poised”, with the transcription machinery in place, ready to start transcription upon removal of the H3K27me3 mark (reviewed in [51]). ESC do not exist in schistosomes, but totipotent somatic stem cells (neoblasts) were found in platyhelminthes from the free-living turbellarians (for review [52]) to the endo- or ecto-parasitic flatworms (monogenea, cestoda, trematoda) [53]. Recently, somatic stem cells were detected in adult schistosomes and called proliferating somatic cells (PSCs) [23,54]. Another type of stem cells, totipotent germinal cells, is present in the intramolluscan stages of *S. mansoni* but they are restricted to a small germinal cell cluster in cercariae [23]. Chromatin structure profiles for these cells have not been studied but it could be that cercariae possess simply a

large number of ESC-like cells. The ultrastructure of cercariae was described in great detail in the past [55,56] and apart from the above-mentioned small population of germinal cells, no stem cells were detected. Stem cells could be identified thanks to the expression of characteristic marker genes such as *vasa*, *piwi*, *tudor* and *nanos*. However, *piwi* and *tudor*-domain containing genes are absent from flukes and tapeworms and no true *vasa* orthologue has been identified in schistosomes [41]. Thus, we sought to exploit rapid replication as another characteristic feature of these cells to identify them in schistosomes.

No DNA replication in cercariae

We used 5-ethynyl-2'-deoxyuridine (EDU) incorporation to detect replication in cercariae and schistosomula. While the latter showed at least one replicating cell per individual in 27% (36 out of 135) after 2 h EDU pulse and 86% (156 out of 181) after 18 h, in none of the cercariae did we detect replication. Consequently, cercariae do not contain large amounts of rapidly replicating ES-like cells or neoblasts that could explain the ChIP-Seq profiles.

Discussion

We describe here the epigenome of *S. mansoni* using four major histone modifications. We analyzed unique sequences, in particular those containing genes, but also the repetitive compartment of the genome.

The enrichment of histone forms around TSS shares similarities with other metazoans, but also shows particularities. Similar to what is found in model organisms, H3K4me3 is enriched at transcription start sites (TSS) and promoter regions of genes with actual or potential transcription [2,5,57–59]. In mammalian genomes, it marks active or poised (when combined with a repressive mark) high CpG-content promoters, while is only present at low CpG-content promoters when the latter are active. Further analysis will show if this is also the case in schistosomes and for what reason some TSS are devoid of H3K4me3. H3K9ac is another mark associated with euchromatin and facilitation of transcription [2–6]. The main enrichment of this mark is often observed between the beginning of promoters and the 5' end of genes, and sometimes in enhancers [2,3,7]. In *S. mansoni* there is a clear difference between cercariae and adults with H3K9ac spanning the gene body only in the latter. H3K9me3 is a landmark of constitutive heterochromatin. In many organisms it is highly enriched at telomeric, pericentromeric and repeat-rich regions, and is responsible for stable large-scale repression [8–11]. In humans, it was shown that this modification is mutually exclusive with H3K4me3 [12]. This is also the case in adult schistosomes (Fig 6B). Nevertheless, H3K9me3 can be present at a very high level over transcribed regions of active mammalian genes, especially 1.5 kb after the TSS, and at a low levels at the promoter [60]. H3K9me3 increases during activation of transcription and is quickly removed afterwards [60]. In our study, association of H3K9me3 with the TSS is only observed in cercariae, where this mark is predominately located upstream of the gene. In *S. mansoni*, a combination of H3K9me3 and H3K9ac allows the prediction of the direction of transcription (always from the H3K9me3 peak to the H3K9ac peak) and it seems that blocks of H3K9me3 separate promoters of genes that are transcribed back-to-back (example shown in Fig 2C). This might be a signature of protein-coding genes. The fourth studied histone isoform, H3K27me3 is known to maintain facultative heterochromatin and is involved in the regulation and sequential expression/repression of developmental genes [11,61]. It is present at inactive or poised (when combined with H3K4me3) high CpG-content promoters and is anti-correlated with the level of transcription [9,10]. In some unicellular organisms, such as *Plasmodium falciparum* this mark can even be absent [6]. In *S. mansoni*, the mark is weak in genes of schistosomula and adults, but present in cercariae. Similar to ESC, bivalent trimethylation of H3K4

and H3K27 is present in at least 121 genes. Bivalent methylation is often considered as an ESC specific mark that leads to “poised” transcription that can rapidly be activated when the cells enter into differentiation. Our data indicate that there are no rapidly proliferating stem cells in cercariae, and we show that there is no transcriptional activity. Consequently, neoblasts of cercariae are either in a dormant state, or somatic cells of cercariae use the same mechanism as ESC to poise transcription that is rapidly activated after infection, upon transformation into schistosomula (at least during the first 18 h after penetration in the host). Bivalent H3K4/H3K27 trimethylation could therefore be a phylogenetically very old and/or more general mechanism to poise transcription than currently thought. However, as only 121 genes were detected to have these bivalent marks, another mechanisms could be involved in the transcription silencing observed in cercariae. At the moment, we can only hypothesize that the major shift of H3K27me3 enrichment (from gene body in cercariae to upstream of TSS in schistosomula and adults, [Fig 4](#)) and/or the strong enrichment of H3K9me3 and H3K9ac at the TSS of cercariae play a role in this phenomenon.

While the importance of repetitive sequences for genome structure and evolution is generally accepted, they are often considered too difficult to analyze by genome-wide approaches. Almost half of the *S. mansoni* genome is composed of repeats [\[39\]](#) and we deemed it to be important to include these in our analysis. Here we used a straightforward approach to evaluate chromatin structure in all known repeat classes of *S. mansoni*. We show that in five mini-satellite regions composed of retrotransposons H3K4me3 and H3K9ac increase while H3K9me3 and H3K27me3 decrease during the transition from cercariae into adults. It is difficult to evaluate the functional importance of this chromatin structure change. It was shown in cultured human cells that another type of satellite DNA (gamma satellites) is enriched for H3K4me3, and that it shields adjacent genes from heterochromatisation [\[62\]](#). Acetylation of H3K9 was not investigated in this latter study. A similar effect might occur in our system but further work will be necessary to elucidate the function of these repeat regions.

In summary, we present here the epigenomes of *S. mansoni* for three life cycle stages. We show that in cercariae, bivalent H3K4/H3K27 trimethylation exists. To the best of our knowledge, this bivalent trimethylation has, in the animal kingdom, so far only been found in vertebrate ESCs. We also show that TSS occupancy by H3K4me3 is stable between cercariae and adults in the large majority of genes. It changes drastically for H3K9 trimethylation and acetylation. In addition, we identified at least five repeat locations that gain euchromatic status during development into adults. Our findings support a model in which transcription is poised in cercariae due to the presence of both activating and repressive marks. It is reactivated, presumably through the removal of the repressive methylation at the TSS and in the gene body, and in conjunction with large scale chromatin remodeling (indicated by changes in repeat regions). In this model, histone-modifying enzymes, in particular H3K27 and H3K9 demethylases would act within 18 h after skin penetration to release the poised state and allow for rapid activation of the transcription machinery that is already in place around the TSS. We had shown before that histone-modifying enzymes are indeed essential for transformation and infection success of another life cycle stage, the miracidia [\[63\]](#). Others have shown that inhibition of HDAC8 leads to parasite death [\[64\]](#). In zebrafish, a H3K27 demethylase of the KDM6 type (UTX) is needed for fin regeneration [\[65\]](#). An orthologue of UTX is encoded in the *S. mansoni* genome (Smp_034000) and by q-RT-PCR we found that mRNA for this gene is present in cercariae, all four stages of schistosomula *in vivo*, and in male and female adults. Further studies will show if blocking this candidate will result in developmental arrest. Besides being of academic interest, such enzymes would also be promising new drug candidates, especially in the light of the fact that the only drug currently available to cure schistosomiasis, praziquantel, is ineffective against the larval forms [\[66\]](#). In addition, many parasites possess infective life cycle stages

that develop immediately after infection. Our findings should motivate search for characteristic histone modifications in these stages.

Supporting Information

S1 Table. (a) Primers targeting regions with bivalent H3K4me3 and H3K27me3 marks in cercariae, and monovalent H3K4me3 marks in adults. (b) Primers targeting regions with H3K4me3 differences between cercariae and adults.

(DOCX)

S2 Table. H3K4me3 peaks in putative polycistrons with a maximum intergenic distance of 200bp. Upstream indicates the transcript that is located upstream of the trans-splicing acceptor site. Down-stream indicates the putative trans-spliced transcript. Polycistron list from <http://www.plosntds.org/article/info%3Adoi%2F10.1371%2Fjournal.pntd.0001455#s5> (S2 Table).

(XLS)

S3 Table. DESeq results for ChIP-Seq on repetitive sequences.

(XLSX)

S4 Table. List of transcripts for which TSS has H3K4me3 and H3K27me in cercariae, evidence from GeneDB (for Smp_#####) and from AmiGO v.1.8 (for TCONS_#####).

(PDF)

S1 Fig. Relative enrichment compared to the alpha-tubulin locus: ChIP-Seq vs qPCR, p-value is given for the t-test of the null hypothesis that the corresponding slope is equal to zero against the two-sided alternative.

(TIFF)

Acknowledgments

The authors are grateful for technical support by Anne Rognon, Jean-François Allienne, Nathalie Aranciabla, Rémi Emans, Bernard Dejean and Claire Picard. We thank Ivonne Heinze for her diligent and skillful technical assistance in Illumina sequencing. We also thank the French National Research Agency for providing financial support to this project (ANR INVIMORY).

Author Contributions

Conceived and designed the experiments: CG. Performed the experiments: DR MJML MALP MF HP MG. Analyzed the data: DR CC CG. Contributed reagents/materials/analysis tools: RE. Wrote the paper: DR MJML MALP CC CG.

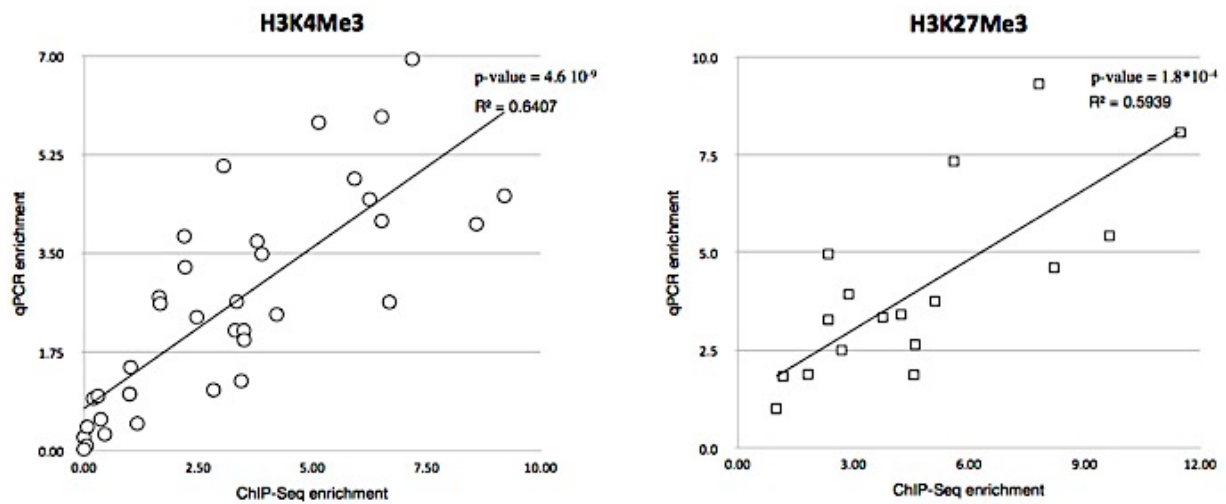
References

1. Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 403: 41–45. PMID: [10638745](#)
2. Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM et al. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122: 517–527. PMID: [16122420](#)
3. Nishida H, Suzuki T, Kondo S, Miura H, Fujimura Y et al. (2006) Histone H3 acetylated at lysine 9 in promoter is associated with low nucleosome density in the vicinity of transcription start site in human cell. *Chromosome Res* 14: 203–211. PMID: [16544193](#)
4. Koch CM, Andrews RM, Flicek P, Dillon SC, Karaoz U et al. (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res* 17: 691–707. PMID: [17567990](#)

5. Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC et al. (2011) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 471: 480–485. doi: [10.1038/nature09725](https://doi.org/10.1038/nature09725) PMID: [21179089](https://pubmed.ncbi.nlm.nih.gov/21179089/)
6. Hoeijmakers WA, Stunnenberg HG, Bartfai R (2012) Placing the *Plasmodium falciparum* epigenome on the map. *Trends Parasitol* 28: 486–495. doi: [10.1016/j.pt.2012.08.006](https://doi.org/10.1016/j.pt.2012.08.006) PMID: [22999479](https://pubmed.ncbi.nlm.nih.gov/22999479/)
7. Hon GC, Hawkins RD, Ren B (2009) Predictive chromatin signatures in the mammalian genome. *Hum Mol Genet* 18: R195–201. doi: [10.1093/hmg/ddp409](https://doi.org/10.1093/hmg/ddp409) PMID: [19808796](https://pubmed.ncbi.nlm.nih.gov/19808796/)
8. Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M et al. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J* 24: 800–812. PMID: [15678104](https://pubmed.ncbi.nlm.nih.gov/15678104/)
9. Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I et al. (2011) Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res* 21: 227–236. doi: [10.1101/gr.115519.110](https://doi.org/10.1101/gr.115519.110) PMID: [21177964](https://pubmed.ncbi.nlm.nih.gov/21177964/)
10. Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12: 7–18. doi: [10.1038/nrg2905](https://doi.org/10.1038/nrg2905) PMID: [21116306](https://pubmed.ncbi.nlm.nih.gov/21116306/)
11. Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21: 381–395. doi: [10.1038/cr.2011.22](https://doi.org/10.1038/cr.2011.22) PMID: [21321607](https://pubmed.ncbi.nlm.nih.gov/21321607/)
12. Binda O, LeRoy G, Bua DJ, Garcia BA, Gozani O et al. (2010) Trimethylation of histone H3 lysine 4 impairs methylation of histone H3 lysine 9: regulation of lysine methyltransferases by physical interaction with their substrates. *Epigenetics* 5: 767–775. doi: [10.4161/epi.5.8.13278](https://doi.org/10.4161/epi.5.8.13278) PMID: [21124070](https://pubmed.ncbi.nlm.nih.gov/21124070/)
13. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ et al. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125: 315–326. PMID: [16630819](https://pubmed.ncbi.nlm.nih.gov/16630819/)
14. Vastenhouw NL, Schier AF (2012) Bivalent histone modifications in early embryogenesis. *Curr Opin Cell Biol* 24: 374–386. doi: [10.1016/j.ceb.2012.03.009](https://doi.org/10.1016/j.ceb.2012.03.009) PMID: [22513113](https://pubmed.ncbi.nlm.nih.gov/22513113/)
15. Hotez PJ, Brindley PJ, Bethony JM, King CH, Pearce EJ et al. (2008) Helminth infections: the great neglected tropical diseases. *J Clin Invest* 118: 1311–1321. doi: [10.1172/JCI34261](https://doi.org/10.1172/JCI34261) PMID: [18382743](https://pubmed.ncbi.nlm.nih.gov/18382743/)
16. Brink LH, McLaren DJ, Smithers SR (1977) *Schistosoma mansoni*: a comparative study of artificially transformed schistosomula and schistosomula recovered after cercarial penetration of isolated skin. *Parasitology* 74: 73–86. PMID: [320543](https://pubmed.ncbi.nlm.nih.gov/320543/)
17. Anderson L, Pierce RJ, Verjovski-Almeida S (2012) *Schistosoma mansoni* histones: from transcription to chromatin regulation; an in silico analysis. *Mol Biochem Parasitol* 3: 105–114.
18. Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M et al. (2012) A Systematically Improved High Quality Genome and Transcriptome of the Human Blood Fluke *Schistosoma mansoni*. *PLoS Negl Trop Dis* 6: e1455. doi: [10.1371/journal.pntd.0001455](https://doi.org/10.1371/journal.pntd.0001455) PMID: [22253936](https://pubmed.ncbi.nlm.nih.gov/22253936/)
19. Duvall RH, DeWitt WB (1967) An improved perfusion technique for recovering adult schistosomes from laboratory animals. *Am J Trop Med Hyg* 16: 483–486. PMID: [4952149](https://pubmed.ncbi.nlm.nih.gov/4952149/)
20. Mann VH, Morales ME, Rinaldi G, Brindley PJ (2010) Culture for genetic manipulation of developmental stages of *Schistosoma mansoni*. *Parasitology* 137: 451–462. doi: [10.1017/S0031182009991211](https://doi.org/10.1017/S0031182009991211) PMID: [19765348](https://pubmed.ncbi.nlm.nih.gov/19765348/)
21. Lazdins JK, Stein MJ, David JR, Sher A (1982) *Schistosoma mansoni*: rapid isolation and purification of schistosomula of different developmental stages by centrifugation on discontinuous density gradients of Percoll. *Exp Parasitol* 53: 39–44. PMID: [6276213](https://pubmed.ncbi.nlm.nih.gov/6276213/)
22. Basch PF (1981) Cultivation of *Schistosoma mansoni* in vitro. I. Establishment of cultures from cercariae and development until pairing. *J Parasitol* 67: 179–185. PMID: [7241277](https://pubmed.ncbi.nlm.nih.gov/7241277/)
23. Wang B, Collins JJr, Newmark PA (2013) Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni*. *Elife* 2: e00768. doi: [10.7554/eLife.00768](https://doi.org/10.7554/eLife.00768) PMID: [23908765](https://pubmed.ncbi.nlm.nih.gov/23908765/)
24. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8: R19. PMID: [17291332](https://pubmed.ncbi.nlm.nih.gov/17291332/)
25. Portela J, Grunau C, Cosseau C, Beltran S, Dantec C et al. (2010) Whole-genome in-silico subtractive hybridization (WISH)—using massive sequencing for the identification of unique and repetitive sex-specific sequences: the example of *Schistosoma mansoni*. *BMC Genomics* 11: 387. doi: [10.1186/1471-2164-11-387](https://doi.org/10.1186/1471-2164-11-387) PMID: [20565937](https://pubmed.ncbi.nlm.nih.gov/20565937/)
26. Cosseau C, Azzi A, Smith K, Freitag M, Mitta G et al. (2009) Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters. *Molecular & Biochemical Parasitology* 166: 70–76.
27. Cosseau C, Grunau C (2011) Native chromatin immunoprecipitation. *Methods Mol Biol* 791: 195–212. doi: [10.1007/978-1-61779-316-5_15](https://doi.org/10.1007/978-1-61779-316-5_15) PMID: [21913081](https://pubmed.ncbi.nlm.nih.gov/21913081/)

28. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L et al. (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* 15: 1451–1455. PMID: [16169926](#)
29. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359. doi: [10.1038/nmeth.1923](#) PMID: [22388286](#)
30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. doi: [10.1093/bioinformatics/btp352](#) PMID: [19505943](#)
31. Althammer S, Gonzalez-Vallinas J, Ballare C, Beato M, Eyra E (2011) Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics* 27: 3333–3340. doi: [10.1093/bioinformatics/btr570](#) PMID: [21994224](#)
32. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842. doi: [10.1093/bioinformatics/btq033](#) PMID: [20110278](#)
33. Feng X, Grossman R, Stein L (2011) PeakRanger: a cloud-enabled peak caller for ChIP-seq data. *BMC Bioinformatics* 12: 139. doi: [10.1186/1471-2105-12-139](#) PMID: [21554709](#)
34. Stein LD, Mungall C, Shu S, Caudy M, Mangone M et al. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res* 12: 1599–1610. PMID: [12368253](#)
35. Trapnell C, Roberts A, Goff L, Pertea G, Kim D et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7: 562–578. doi: [10.1038/nprot.2012.016](#) PMID: [22383036](#)
36. Stark R, Brown G (2013) DiffBind: differential binding analysis of ChIP-Seq peak data. *bioconductorjrp*
37. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80. PMID: [15461798](#)
38. Hebenstreit D, Gu M, Haider S, Turner DJ, Lio P et al. (2011) EpiChIP: gene-by-gene quantification of epigenetic modification levels. *Nucleic Acids Res* 39: e27. doi: [10.1093/nar/gkq1226](#) PMID: [21131282](#)
39. Lepesant JM, Roquis D, Emans R, Cosseau C, Arancibia N et al. (2012) Combination of de novo assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in *Schistosoma mansoni*. *Exp Parasitol* 130: 470–474. doi: [10.1016/j.exppara.2012.02.010](#) PMID: [22381218](#)
40. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106. doi: [10.1186/gb-2010-11-10-r106](#) PMID: [20979621](#)
41. Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A et al. (2013) The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* 496: 57–63. doi: [10.1038/nature12031](#) PMID: [23485966](#)
42. Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP et al. (2009) The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460: 352–358. doi: [10.1038/nature08160](#) PMID: [19606141](#)
43. Gao D, Kim J, Kim H, Phang TL, Selby H et al. (2010) A survey of statistical software for analysing RNA-seq data. *Hum Genomics* 5: 56–60. PMID: [21106489](#)
44. Lepesant JM, Cosseau C, Boissier J, Freitag M, Portela J et al. (2012) Chromatin structure changes around satellite repeats on the *Schistosoma mansoni* female sex chromosome suggest a possible mechanism for sex chromosome emergence. *Genome Biol* 13: R14. doi: [10.1186/gb-2012-13-2-r14](#) PMID: [22377319](#)
45. Grossman AI, Short RB, Cain GD (1981) Karyotype evolution and sex chromosome differentiation in Schistosomes (Trematoda, Schistosomatidae). *Chromosoma* 84: 413–430. PMID: [7327052](#)
46. Chueh AC, Wong LH, Wong N, Choo KH (2005) Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. *Hum Mol Genet* 14: 85–93. PMID: [15537667](#)
47. Mourao MM, Grunau C, Loverde PT, Jones MK, Oliveira G (2012) Recent advances in *Schistosoma* genomics. *Parasite Immunol* 34: 151–162. doi: [10.1111/j.1365-3024.2011.01349.x](#) PMID: [22145587](#)
48. Fitzpatrick JM, Peak E, Perally S, Chalmers IW, Barrett J et al. (2009) Anti-schistosomal intervention targets identified by lifecycle transcriptomic analyses. *PLoS Negl Trop Dis* 3: e543. doi: [10.1371/journal.pntd.0000543](#) PMID: [19885392](#)
49. Theron A, Pointier J-P (1985) Recherche des facteurs susceptibles d'empêcher la réalisation du cycle de *Schistosoma mansoni* dans les mares de la grande terre de Guadeloupe (Antilles françaises). *Annales de parasitologie humaine et comparée* 60:
50. Pan G, Tian S, Nie J, Yang C, Ruotti V et al. (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1: 299–312. doi: [10.1016/j.stem.2007.08.003](#) PMID: [18371364](#)

51. Vastenhouw NL, Zhang Y, Woods IG, Imam F, Regev A et al. (2010) Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* 464: 922–926. doi: [10.1038/nature08866](https://doi.org/10.1038/nature08866) PMID: [20336069](https://pubmed.ncbi.nlm.nih.gov/20336069/)
52. Baguna J (2012) The planarian neoblast: the rambling history of its origin and some current black boxes. *Int J Dev Biol* 56: 19–37. doi: [10.1387/ijdb.113463jb](https://doi.org/10.1387/ijdb.113463jb) PMID: [22252540](https://pubmed.ncbi.nlm.nih.gov/22252540/)
53. Brehm K (2010) *Echinococcus multilocularis* as an experimental model in stem cell research and molecular host-parasite interaction. *Parasitology* 137: 537–555. doi: [10.1017/S0031182009991727](https://doi.org/10.1017/S0031182009991727) PMID: [19961652](https://pubmed.ncbi.nlm.nih.gov/19961652/)
54. Collins JJr, Newmark PA (2013) It's no fluke: the planarian as a model for understanding schistosomes. *PLoS Pathog* 9: e1003396. doi: [10.1371/journal.ppat.1003396](https://doi.org/10.1371/journal.ppat.1003396) PMID: [23874195](https://pubmed.ncbi.nlm.nih.gov/23874195/)
55. Collins JJr, Wang B, Lambrus BG, Tharp ME, Iyer H et al. (2013) Adult somatic stem cells in the human parasite *Schistosoma mansoni*. *Nature* 494: 476–479. doi: [10.1038/nature11924](https://doi.org/10.1038/nature11924) PMID: [23426263](https://pubmed.ncbi.nlm.nih.gov/23426263/)
56. Dorsey CH, Cousin CE, Lewis FA, Stirewalt MA (2002) Ultrastructure of the *Schistosoma mansoni* cercaria. *Micron* 33: 279–323. PMID: [11742750](https://pubmed.ncbi.nlm.nih.gov/11742750/)
57. Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature* 447: 407–412. PMID: [17522673](https://pubmed.ncbi.nlm.nih.gov/17522673/)
58. Gu SG, Fire A (2010) Partitioning the *C. elegans* genome by nucleosome modification, occupancy, and positioning. *Chromosoma* 119: 73–87. doi: [10.1007/s00412-009-0235-3](https://doi.org/10.1007/s00412-009-0235-3) PMID: [19705140](https://pubmed.ncbi.nlm.nih.gov/19705140/)
59. Zentner GE, Henikoff S (2013) Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* 20: 259–266. doi: [10.1038/nsmb.2470](https://doi.org/10.1038/nsmb.2470) PMID: [23463310](https://pubmed.ncbi.nlm.nih.gov/23463310/)
60. Vakoc CR, Mandat SA, Olenchok BA, Blobel GA (2005) Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol Cell* 19: 381–391. PMID: [16061184](https://pubmed.ncbi.nlm.nih.gov/16061184/)
61. Trojer P, Reinberg D (2007) Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell*
62. Kim JH, Ebersole T, Kouprina N, Noskov VN, Ohzeki J et al. (2009) Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. *Genome Res* 19: 533–544. doi: [10.1101/gr.086496.108](https://doi.org/10.1101/gr.086496.108) PMID: [19141594](https://pubmed.ncbi.nlm.nih.gov/19141594/)
63. Azzi A, Cosseau C, Grunau C (2009) *Schistosoma mansoni*: developmental arrest of miracidia treated with histone deacetylase inhibitors. *Exp Parasitol* 121: 288–291. doi: [10.1016/j.exppara.2008.11.010](https://doi.org/10.1016/j.exppara.2008.11.010) PMID: [19094990](https://pubmed.ncbi.nlm.nih.gov/19094990/)
64. Marek M, Kannan S, Hauser AT, Moraes Mourao M, Caby S et al. (2013) Structural basis for the inhibition of histone deacetylase 8 (HDAC8), a key epigenetic player in the blood fluke *Schistosoma mansoni*. *PLoS Pathog* 9: e1003645. doi: [10.1371/journal.ppat.1003645](https://doi.org/10.1371/journal.ppat.1003645) PMID: [24086136](https://pubmed.ncbi.nlm.nih.gov/24086136/)
65. Stewart S, Tsun ZY, Izpisua Belmonte JC (2009) A histone demethylase is necessary for regeneration in zebrafish. *Proc Natl Acad Sci U S A* 106: 19889–19894. doi: [10.1073/pnas.0904132106](https://doi.org/10.1073/pnas.0904132106) PMID: [19897725](https://pubmed.ncbi.nlm.nih.gov/19897725/)
66. Doenhoff MJ, Cioli D, Utzinger J (2008) Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Curr Opin Infect Dis* 21: 659–667. doi: [10.1097/QCO.0b013e328318978f](https://doi.org/10.1097/QCO.0b013e328318978f) PMID: [18978535](https://pubmed.ncbi.nlm.nih.gov/18978535/)



Supplementary Figure 1 : relative enrichment compared to the alpha-tubulin locus : ChIP-Seq vs. qPCR. P-value is given for the t-test of the null hypothesis that the corresponding slope is equal to zero against the two-sided alternative.

Supplementary Table 1:

(A) Primers targeting regions with bivalent H3K4me3 and H3K27me3 marks in cercaria, and monovalent H3K4me3 marks in adults

Name	Forward sequence	Reverse sequence
sm.Chr_1:47971550,47972100	gcacgtaaaccaccagcccca	caagcgaaaatccacgcgcca
sm.SC_0034:2246640,2247240	gctcgcggagagtaaaccgga	acgccgcctgaagatcaaggg
sm.SC_0071:1183830,1184320	tgaccagtgtggaaatgacctgtgt	agtgaccactcgataggcacgca
sm.Chr_1:8165260,8166110	gacctgggtgcactgggcaa	tggacgaaagtctcggttctgatctct
sm.Chr_2:21540210,21540530	tgctttgtgtgattccaaagcccatt	agatcacagcccgagccttct
sm.SC_0225:87490,88050	tgtccagcaaaatcccatgtcgc	tgtgttgctgtgtgccacagga
sm.Chr_2:26997540,26998000	tcgtctcgcaattacttttgagcca	caatgactgtgttgcggaatatctggt
sm.Chr_3:12856600,12856890	tcctccacccccacgtacgttat	ccggttgagttgccatacccca
sm.Chr_1:40471080,40471790	tggttagcttacgtgcgagagttc	acgcgcttgtagtgtgactgt
sm.Chr_W:44511380,44511690	tgaggcgagataggacacattcattca	agatggtcattaggaacagccgaact

(B) Primers targeting regions with H3K4me3 differences between cercaria and adults

Name	Forward sequence	Reverse sequence
XLOC_002337	ccgtcacggccaataagggcg	gctcccatatgatgctcgattggct
XLOC_004599	tgagcgagagtgatgctcttttcg	acagaagctgacgcccctga
XLOC_004718	aagtggacgagaccggccca	aggttgcgaaaagccgacgaca
XLOC_010397	tgggtggtggttgctgtgga	cgccccaaaactcgcaagcc
XLOC_004811	gcctgacgctagcgactgaaca	tccggacaggttggggtgtct
XLOC_011635	agaaccactgaggggtatggtgtgt	ttcgctcctccagcccacaa
XLOC_001136	gggcggcagaatcagggtcg	gggaactccggagaccaca
XLOC_006515	tccaccactttcaccgaccact	acgtcagctgtcgctactacca
XLOC_006717	gctactcaacatatcctgatctcccg	gattgagctctattttgaggaagccga
XLOC_001871	ccgatcgtgcgcttcgtcgat	agccttgatgccagctttccga

Supplementary Table 2: (A) Putative polycistrons with a maximum intergenic distance of 200bp. Up-stream indicates the transcript that is located up-stream of the trans-splicing acceptor site. Down-stream indicates the putative trans-spliced transcript. From <http://www.plosntds.org/article/info%3Adoi%2F10.1371%2Fjournal.pntd.0001455#s5> (Table S2)

Up-stream	Down-stream	H3K4me3 peak	Genes	Comments
Smp_034530.1	Smp_140830.1	Smp_034530.1	2	
Smp_165510.1	Smp_129050.1	Smp_129050.1	2	Smp_165510.1 intragenic
Smp_136210.1	Smp_025950.1	both	2	
Smp_025680.1	Smp_025670.1	Smp_025680.1	2	
Smp_057140.1	Smp_210460	Smp_057140.1	2	in red: corrected
Smp_021280.1	Smp_133700.1	Smp_021280.1	2	
Smp_187170.1	Smp_122050.1	no peak		
Smp_048870.1	Smp_048880.1	Smp_048870.1	2	
Smp_023160.1	Smp_023170.1	double peak	2	
Smp_180990.1	Smp_181000.1	Smp_180990.1	2	
Smp_033590.1	Smp_140300.1	6 kb upstream of Smp_033590.1	3	
Smp_079750.1	Smp_079760.1	Smp_079750.1	2	
Smp_084900.1	Smp_084890.1	Smp_084900.1	2	
Smp_030030.1	Smp_030020.1	Smp_030030.1	2	
Smp_120980.1	Smp_120960.1	Smp_120980.1	2 or 3	Smp_120980 intragenic. Tiny peak of H3K4me3 at the start of Smp_120960. 1 additional "gene" in our transcriptome
Smp_114470.1	Smp_138100.1	Smp_138100.1	2 to 4	Smp_114470 intragenic, which is huge (80 kb). 2 additional "genes" in our transcriptome
Smp_006980.1	Smp_006970.1	Smp_006980.1	2	
Smp_188210.1	Smp_172620.1	inside Smp_172620	2	Smp_188210 intragenic
Smp_146660.1	Smp_211050	Upstream of smp_146660.1	2 or 3	in red: corrected
Smp_038630.1	Smp_038640.1	Smp_038630.1	2	
Smp_058970.1	Smp_153860.2	Smp_058970.1	2	
Smp_090790.1	Smp_090800.1	Smp_090790.1	2	
Smp_090830.1	Smp_090820.1	Upstream of Smp_090830	2	
Smp_175220.1	Smp_175210.1	Smp_175220.1	2 or 3	Smp_175200 downstream of Smp_175210 do not have a H3K4me3 mark
Smp_149120.1	Smp_149100.1	Inside Smp_149120	2 to 4	Smp_149100, additional "genes" in our transcriptome
Smp_101220.1	Smp_127110.1	Smp_127110.1	2	Smp_101220 intragenic
Smp_191360.1	Smp_157000.1	Smp_157000.1	2	Smp_191360 intragenic
Smp_143000.1	Smp_142990.1	Smp_143000.1	2	
Smp_038430.1	Smp_038420.1	Smp_038430.1	3	
Smp_114910.1	Smp_171960.1	Inside or upstream of Smp_171960	2	Smp_114910 intragenic
Smp_165540.1	Smp_080520.1	Upstream of Smp_165540	2	
Smp_128610.1	Smp_012050.1	Smp_128610.1	2	
Smp_088390.1	Smp_088380.1	Upstream of Smp_088390.1	3	
Smp_088380.1	Smp_170260.1	Upstream of Smp_088390.1	3	
Smp_059870.1	Smp_059880.1	Smp_059870.1	2	
Smp_173910.1	Smp_094470.1	Smp_173910.1	2	
Smp_024120.1	Smp_024110.1	Upstream of Smp_024120.1	2	very wide H3K4me3 mark
Smp_180370.1	Smp_180360.1	Smp_180370.1	2	
Smp_162350.1	Smp_196030.1	Smp_162350.1	2	
Smp_162910.1	Smp_192920.1	Smp_192920.1	2	Smp_192920 intragenic
Smp_160260.1	Smp_160270.1	Smp_160260.1	2	
Smp_180190.1	Smp_133910.1	Smp_133910.1	1	No sign of Smp_133910
Smp_103330.1	Smp_103320.1	Upstream of Smp_103330.1	2	
Smp_155580.1	Smp_155560.1	Smp_155560.1	1	Both genes very far appart with a gene in the reverse orientation between them
Smp_041680.1	Smp_041680.1	Upstream of Smp_041680.1	2	
Smp_104530	Smp_180530.4	Smp_180530.4	2	in red: corrected

(B) Putative polycistrons with a maximum intergenic distance of 2000 bp. Up-stream indicates the transcript that is located up-stream of the trans-spliced acceptor site.

Up-stream	Down-stream	H3K4me3 peak	Genes	Comments
Smp_003770.1	Smp_003760.1	wide	2	
Smp_003630.1	Smp_124110.1	Smp_003630.1	2	
Smp_035010.1	Smp_035000.1	no peak	3	Upstream gene Smp_035020.1 display a peak and might be the first gene in the polycystronic unit. Transcriptomic evidences support this hypthothesis.
Smp_020370.1	Smp_020360.1	Smp_020370.1	2 or more	
Smp_020210.1	Smp_133160.1	Smp_020210.1	2	
Smp_034530.1	Smp_140830.1	Smp_034530.1	2	
Smp_012450.1	Smp_012460.1	no peak	2	
Smp_165510.1	Smp_129050.1	Smp_165510.1	2	Smp_129050.1 intragenic
Smp_034030.1	Smp_140550.1	no peak	2 or 3	Upstream gene Smp_034010.1 display a peak and might be the first gene in the polycystronic unit. Transcriptomic evidences support this hypthothesis.
Smp_200830.1	Smp_200840.1	no peak	2 or more	
Smp_136210.1	Smp_025950.1	Upstream of Smp_136210.1	2	
Smp_025680.1	Smp_025670.1	Smp_025680.1	2	
Smp_066820.1	Smp_066830.1	Smp_066820.1	2	
Smp_114160.1	Smp_196480.1	Upstream of Smp_114160.1	2	
Smp_152750.1	Smp_152760.1	Smp_152750.1	2	
Smp_057140.1	Smp_152820.1	Smp_057140.1	2	
Smp_165030.1	Smp_079570.1	Smp_165030.1	2	
Smp_173050.2	Smp_092990.1	Smp_173050.2	2	
Smp_021280.1	Smp_133700.1	Smp_021280.1	2	
Smp_147940.1	Smp_147930.1	Smp_147940.1	2	
Smp_187170.1	Smp_122050.1	no peak	2	Smp_187170.1 intragenic
Smp_127340.1	Smp_127350.1	Smp_127340.1	2	
Smp_038820.1	Smp_143290.1	Upstream of Smp_038820.1	2 or 3	
Smp_176620.1	Smp_099240.1	Smp_176620.1	2	
Smp_121610.3	Smp_194560.1	Smp_121610.3	2	
Smp_048870.1	Smp_048880.1	Smp_048870.1	2	
Smp_049000.1	Smp_048990.1	no peak	2	
Smp_012410.1	Smp_012400.1	Smp_012410.1	2 or 3	
Smp_023160.1	Smp_023170.1	Smp_023160.1	2	

Smp_180990.1	Smp_181000.1	Smp_180990.1	2	
Smp_033590.1	Smp_140300.1	Upstream of Smp_033590.1	3	
Smp_140400.1	Smp_033770.1	Smp_140400.1	2	
Smp_075030.1	Smp_162650.1	Smp_075030.1	2	
Smp_074930.1	Smp_074940.1	Smp_074930.1	2	
Smp_150910.1	Smp_160570.1	Smp_150910.1	2 or more	Smp_150910.1 intragenic
Smp_026250.3	Smp_026260.1	Smp_026250.3	2 or 3	
Smp_195160.1	Smp_076650.1	Smp_076650.1	2	
Smp_123620.1	Smp_123610.1	Smp_123620.1	2	
Smp_123530.1	Smp_123520.1	Smp_123530.1	2	
Smp_079750.1	Smp_079760.1	Smp_079750.1	2	
Smp_084900.1	Smp_084890.1	Smp_084900.1	2	
Smp_030030.1	Smp_030020.1	Smp_030030.1	2	
Smp_120980.1	Smp_120960.1	Smp_120980.1	2	Smp_120980.1 intragenic
Smp_042690.1	Smp_042680.1	Smp_042690.1	2	
Smp_114470.1	Smp_138100.1	Smp_138100.1	2	Smp_114470.1 intragenic
Smp_006980.1	Smp_006970.1	Smp_006980.1	2	
Smp_188210.1	Smp_172620.1	Upstream of Smp_188210.1	2	Smp_188210.1 intragenic
Smp_150050.1	Smp_150040.2	inside Smp_150050.1	2	
Smp_146660.1	Smp_146670.1	upstream of Smp_146660.1	2	
Smp_038630.1	Smp_038640.1	Upstream of Smp_038630.1	2	
Smp_066960.2	Smp_066970.1	Smp_066960.2	2	
Smp_058970.1	Smp_153860.2	Smp_058970.1	2	
Smp_090790.1	Smp_090800.1	Smp_090790.1	2	
Smp_090830.1	Smp_090820.1	no peak	2	Upstream gene Smp_171770.1 display a peak and might be the first gene in the polycystronic unit. Transcriptomic evidences support this hypothesis.
Smp_070170.1	Smp_070200.1	Smp_070170.1	2	
Smp_175220.1	Smp_175210.1	Smp_175220.1	2 or 3	
Smp_042420.1	Smp_042430.1	Smp_042420.1	2	
Smp_080130.1	Smp_165320.1	Smp_080130.1	2 or more	
Smp_149120.1	Smp_149100.1	downstream of Smp_149100.1	2	Smp_149100.1 intragenic
Smp_027370.1	Smp_027360.1	no peak	2 or 3	
Smp_035730.1	Smp_141410.1	Smp_035730.1	2	
Smp_089060.2	Smp_089050.1	Smp_089060.2	2	
Smp_126750.1	Smp_126760.1	Smp_126750.1	2	
Smp_049930.1	Smp_049920.1	Smp_049930.1	2	
Smp_064400.1	Smp_064390.1	no peak	2	
Smp_155210.1	Smp_155200.1	Smp_155210.1	2	
Smp_052760.1	Smp_052750.1	Smp_052760.1	2	
Smp_024930.1	Smp_135720.1	10 kb upstream of Smp_024930.1	2	Transcriptomic data shows additional upstream exons for Smp_024930.1 where the peak is located
Smp_149360.1	Smp_149350.1	Smp_149360.1	2	
Smp_127110.1	Smp_101220.1	Smp_127110.1	2	Smp_101220 intragenic, order changed
Smp_133850.1	Smp_133860.1	Smp_1021550	3	Smp_021550 is the gene upstream of Smp_133850
Smp_157000.1	Smp_191360.1	Smp_157000	2 or 3	Smp_191360 intragenic, order changed, additional peak + transcript within predicted Smp_157000
Smp_143000.1	Smp_142990.1	Smp_143000.1	2	additional peak within Smp_142990
Smp_038430.1	Smp_038410.1	Smp_038430.1	3	additional gene upstream of originally identified Smp_038420
Smp_038430.1	Smp_038420.1			fusionated with line above
Smp_036630.1	Smp_141890.1	Smp_036630.1	2	
Smp_146490.1	Smp_045060.1	Smp_146490.1	2	
Smp_032920.2	Smp_032930.1	Smp_032920.2	2	
Smp_171960.1	Smp_114910.1	Smp_171960.1	2	Smp_114910 intragenic, , small peak at Smp_114910
Smp_165540.1	Smp_080520.1	Smp_165540.1	2	
Smp_128610.1	Smp_012050.1	both	2	very small genes, peak covers body of the gene
Smp_088390.1	Smp_170260.1	Smp_088390.1	3	3 instead of predicted 2
Smp_088380.1	Smp_170260.1			fusionated with line above
Smp_059870.1	Smp_059880.1	Smp_059870.1	2	
Smp_165830.1	Smp_165840.2	Smp_165830.1	2	
Smp_173910.1	Smp_094470.1	Smp_173910.1	2	
Smp_131960.1	Smp_131950.1	Smp_131960.1	2	
Smp_018020.1	Smp_018030.1	Smp_018020.1	2	
Smp_066010.2	Smp_157810.1	Smp_066010.2	2 or 3	small peak in 5' of Smp_157810 in adults, downstream genes Smp_157820 without peak in TSS but intragenic
Smp_157710.1	Smp_157700.1	Smp_157710.1	2	
Smp_045960.1	Smp_146870.1	Smp_045960.1	2	
Smp_046090.2	Smp_146900.1	Smp_046090.2	2	
Smp_046120.1	Smp_146910.1	no peak	2	
Smp_024120.1	Smp_024110.1	Smp_024120.1	2	Peak is very wide
Smp_180370.1	Smp_180360.1	Smp_180370.1	2	
Smp_162350.1	Smp_196030.1	Smp_162350.1	2	
Smp_194930.1	Smp_194920.1	Smp_194930.1	2	
Smp_037400.1	Smp_142350.1	Smp_037400.1	2 or 3	Additional gene could be Smp_142340
Smp_015440.1	Smp_130610.1	Upstream of Smp_015440.1	2	
Smp_174990.1	Smp_174980.1	Smp_174990.1	2	
Smp_136530.1	Smp_136520.1	Smp_136530.1	2	
Smp_162910.1	Smp_192920.1	Smp_192920.1	2	Smp_192920.1 intragenic
Smp_160260.1	Smp_160270.1	Smp_160260.1	2	
Smp_077400.1	Smp_077390.1	Smp_077400.1	2	Both genes 2 kb appart
Smp_136580.1	Smp_026700.1	Smp_136580.1	2	
Smp_198750.1	Smp_097710.1	Smp_198750.1	2	
Smp_000640.1	Smp_000630.1	Smp_000640.1	2 or 3	Additional gene could be Smp_000620
Smp_180190.1	Smp_133910.1	Smp_180190.1	1	no sign of Smp_133910.1
Smp_103330.1	Smp_103320.1	Smp_103330.1	2	Wide peak starting upstream of Smp_103330.1
Smp_155580.1	Smp_155560.1	no peak	1	Smp_155560.1 is 34 kb downstream of Smp_155580.1 and Smp_155570.1 is between these two genes and transcribed on the other strand.
Smp_041680.1	Smp_144870.1	Upstream of Smp_041680.1	2	
Smp_042230.1	Smp_145070.1	no peak	2	At the far end of the scaffold
Smp_105070.1	Smp_105080.1	Smp_105070.1	2	
Smp_107240.1	Smp_180530.4	Smp_180530.4	2 or 3	No sign of Smp_107240.1, but Smp_104530 and Smp_180540 are in the vicinity
Smp_164650.1	Smp_078900.1	Smp_164650.1	2	

CHAPITRE III

RÉSISTANCE À UN SCHISTOSOMICIDE ET IMPACT SUR L'ÉPIGÉNOME DE S. MANSONI

III.1. MISE EN CONTEXTE

L'induction de résistance métastable au schistosomicide hycanthone, décrite une première fois par Jansma *et al.* (1977) correspond, à ma connaissance, à la première démonstration d'une *Dauermodifikation* chez *Schistotoma mansoni*. Dans ses expériences, Jansma décrit différents types de traitements sur des souris infectées, impliquant généralement des doses sous-curatives d'hycanthone (*i.e.* insuffisantes pour tuer la majorité des schistosomes), et provoquant l'apparition d'une résistance à la molécule chez les vers adultes survivants et leur descendance. J'étais particulièrement sceptique lors de la lecture de cet article, et il me semblait presque évident que le phénomène observé ne reflétait qu'une sélection des individus naturellement plus tolérants ou résistants au traitement. Toutefois, la métastabilité de cette résistance me laissait quelque peu perplexe. Sur certaines de ses expériences d'induction, Jansma a suivi la descendance de plusieurs de ses lignées induites sur 10 à 21 générations, et le pourcentage de résistance d'une génération à une autre pouvait être hautement fluctuant. Dans un cas particulier, on voyait en l'espace de trois générations, le nombre d'individus résistants passer dans la population de 76%, à 100%, puis 31%. Dans le cas de la sélection d'un gène de résistance, cette dernière devrait rester stable, ou tout du moins suivre une logique d'hérédité mendélienne. La méthodologie de l'article, que j'ai décortiquée plusieurs fois, contient malheureusement plusieurs zones d'ombres qui ne me permettaient pas de trancher hors de tous doutes raisonnables sur l'origine de ce phénomène : véritable *Dauermodifikation* ou résultat de biais expérimentaux?

« Dans le doute, abstiens toi », conseille l'adage. Cette justification ne convenant toutefois pas à mes encadrants, je me suis mis à la recherche de plus d'informations sur cette induction de résistance. J'ai essayé de contacter les auteurs de l'article pour obtenir des précisions sur la méthodologie, mais sans succès. W.B. Jansma n'a publié qu'une seule fois après son article sur l'hycanthone, et a disparu du milieu scientifique. Ernest Bueding, dernier auteur de l'article et parasitologue renommé de l'université John Hopkins de Baltimore est décédé en 1986. J'ai réussi à contacter Steffen H Roger (entre temps devenu président de la Bucknell University en Pennsylvanie), un des co-auteurs, maintenant à la retraite, mais il n'a pas donné suite à ma correspondance. Le dernier signataire de l'article, C.L. Liu, avait un nom beaucoup trop courant pour espérer retrouver sa trace jusqu'à

aujourd'hui. J'ai ensuite fait une recherche bibliographique exhaustive de la résistance à l'hycanthone chez *S. mansoni*. Le traitement par cette molécule ayant été abandonné assez rapidement pour cause de toxicité chez l'homme, la littérature ne comporte qu'une quarantaine d'article sur cette question. Le statut de l'induction de résistance était controversé, une équipe pouvant reproduire les résultats de Jansma, une autre totalement incapable (et décrivant une résistance monogénique récessive et héritable de manière mendélienne). L'explication génétique me semblait bien plus parcimonieuse et convaincante, mais j'ai fait un effort (à l'époque assez important) d'ouverture d'esprit et décidé d'explorer un peu plus en profondeur le phénomène d'induction de résistance.

Durant son doctorat, Julie Lepesant a reproduit une des expériences de Jansma, l'induction de type I (Lepesant 2012). Des souris infectées par des couples de schistosomes ont été injectées avec des doses sous-curatives d'hycanthone, et les œufs des parasites survivants ont été récupérés 60 jours plus tard pour continuer le cycle de vie et tester la résistance à l'état adulte. Julie a utilisé des cercaires issues de l'infection mono-miracidiale de mollusques pour infecter des souris. Pour rappel, toutes les cercaires issues d'un même miracidium sont génétiquement identiques. Ainsi, plusieurs souris sont infectées avec un même génotype du parasite et peuvent servir de réplicats. Cette expérience a été reproduite cinq fois, à partir de cinq infections mono-miracidiales différents (donc cinq génotypes différents du parasite). Julie a observé l'apparition de quelques individus résistants au sein d'une population majoritairement sensible. Dans un contexte où les individus sensibles et résistants sont, en théorie, génétiquement identiques, les modifications épigénétiques étaient une fois de plus de bons candidats pour expliquer les phénotypes. Julie a pu effectuer les expériences de ChIP-Seq sur quatre modifications d'histones sur schistosomes adultes sensibles et résistants. Les expériences n'ont par la suite pas pu être continuées à cause de l'arrêt de la commercialisation de l'hycanthone.

Mon rôle dans ce projet a été de traiter les données de ChIP-Seq avec une approche comparative, pour faire ressortir les différences épigénétiques associées avec l'induction de résistance à l'hycanthone. Entre temps, un gène de résistance à l'hycanthone ayant été identifié (Valentim *et al.* 2013), j'ai effectué des analyses génétiques permettant de montrer qu'il n'était pas présent chez nos individus résistants.

III.2. PUBLICATION ASSOCIÉE

Ce travail a été publié sous le titre « *Exposure to hycanthone alters chromatin structure around specific gene functions and specific repeats in Schistosoma mansoni* » dans le numéro spécial « *Gene Functions in Schistosomes* » de la revue **Frontiers in Genetics**, en août 2014. La *Supplementary Figure 1* et les *Supplementary Tables 1* et *2* de l'article sont présentés à sa suite.



Exposure to hycanthone alters chromatin structure around specific gene functions and specific repeats in *Schistosoma mansoni*

David Roquis^{1,2}, Julie M. J. Lepesant^{1,2}, Emanuel Villafan^{3,4}, Jérôme Boissier^{1,4}, Cristina Vieira^{3,5}, Céline Cosseau^{1,2} and Christoph Grunau^{1,2*}

¹ Département de Biologie, Université de Perpignan Via Domitia, Perpignan, France

² CNRS, UMR 5244, Écologie et Évolution des Interactions (2EI), Perpignan, France

³ CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, Département de Biologie, Université Lyon 1, Villeurbanne, France

⁴ Red de Estudios Moleculares Avanzados, Instituto de Ecología A.C., Xalapa, México

⁵ Département de Biologie, Institut Universitaire de France, Paris, France

Edited by:

Paul J. Brindley, George Washington University, USA

Reviewed by:

Gabriel Rinaldi, George Washington University, USA

Pablo Smircich, Universidad de la Republica, Uruguay

*Correspondence:

Christoph Grunau, Département de Biologie, Université de Perpignan Via Domitia, Perpignan, France
e-mail: christoph.grunau@univ-perp.fr

Schistosoma mansoni is a parasitic plathyhelminth responsible for intestinal schistosomiasis (or bilharzia), a disease affecting 67 million people worldwide and causing an important economic burden. The schistosomicides hycanthone, and its later proxy oxamniquine, were widely used for treatments in endemic areas during the twentieth century. Recently, the mechanism of action, as well as the genetic origin of a stably and Mendelian inherited resistance for both drugs was elucidated in two strains. However, several observations suggested early on that alternative mechanisms might exist, by which resistance could be induced for these two drugs in sensitive lines of schistosomes. This induced resistance appeared rapidly, within the first generation, but was metastable (not stably inherited). Epigenetic inheritance could explain such a phenomenon and we therefore re-analyzed the historical data with our current knowledge of epigenetics. In addition, we performed new experiments such as ChIP-seq on hycanthone treated worms. We found distinct chromatin structure changes between sensitive worms and induced resistant worms from the same strain. No specific pathway was discovered, but genes in which chromatin structure modifications were observed are mostly associated with transport and catabolism, which makes sense in the context of the elimination of the drug. Specific differences were observed in the repetitive compartment of the genome. We finally describe what types of experiments are needed to understand the complexity of heritability that can be based on genetic and/or epigenetic mechanisms for drug resistance in schistosomes.

Keywords: *Schistosoma mansoni*, resistance, induction, hycanthone, oxamniquine, epigenetics, repetitive sequences

INTRODUCTION

Schistosoma mansoni, a platyhelminth trematode parasite, is the causing agent of intestinal schistosomiasis (or bilharzia). The parasite has a heteroxenous life cycle with two hosts and three main developmental stages. The cycle begins when eggs come in contact with fresh water and hatch, releasing a free-swimming larva known as miracidium. This larva infects the intermediate host (a freshwater snail, usually of the *Biomphalaria* genus). Inside the mollusk, it will transform into primary and then secondary sporocysts and undergo asexual multiplication. A second type of free-swimming larvae, cercariae, is released and infects the definitive mammalian host (human or rodent) in which they will reach their adult form (worm) and perform sexual reproduction. Sexually mature schistosomes couples are located in mesenteric veins of the mammalian host. Females lay eggs with a characteristic lateral spine, which allow eggs to go through the wall of the intestine and be excreted in the feces. However, a

large proportion of eggs is caught in the blood circulation and end up being trapped in the liver, leading to hepatic fibrosis. Bilharzia is an endemic disease in many areas in Africa and South America, infecting 67 million people and causing an important socio-economical burden (King, 2010). Since the beginning of the twentieth century, several schistosomicides were developed (Galdino da Rocha Pitta et al., 2013). Many of them were abandoned due to low treatment efficiency, as well as toxic and sometimes mutagenic side effects to the human host (Galdino da Rocha Pitta et al., 2013). Another major concern regarding schistosomicides is the rapid appearance of resistances, both in laboratory strains and natural populations of the parasite (Bruce et al., 1987; Coles et al., 1987; Drescher et al., 1993; WHO, 1998). Please note that to avoid any language confusion, we use here the definitions of drug resistance and tolerance in schistosomes as given by Coles (2002), with the former being “when any isolate has a significantly lower cure rate than the most sensitive isolate” and the

later when the whole species is non-responsive to the drug. These factors left a single drug, praziquantel, as the current therapeutic alternative. As there is very little advance in the development of novel drugs against schistosomiasis, it becomes crucial to deepen our understanding on which molecular and evolutionary mechanisms can lead to schistosomicides resistance. Such knowledge could allow to either avoid the rapid appearance of resistances previously seen in *S. mansoni*, or to find ways to circumvent it.

Various drug resistances in schistosomes were studied through the past decades, but one of the best documented, but also most intriguing one, is against hycanthone. This molecule is not used anymore in treatments, and has been commercially discontinued, but was subjected to many research works from different laboratories. First evidences of resistant worms to this molecule were published in the early 1970's (Rogers and Bueding, 1971). It was later demonstrated several times that schistosomes were in general cross-resistant or cross-susceptible to both hycanthone and another schistosomicide, oxamniquine (Jansma et al., 1977; Cioli and Pica-Mattoccia, 1984; Dias and Olivier, 1985; Bruce et al., 1987). Oxamniquine was often used as a proxy for hycanthone, as the later is not commercially produced anymore and was toxic to the host to a certain degree. Two laboratories obtained stable resistant lines of schistosomes by injecting infected mice with a single curative dose (i.e., killing >90% of the worms in susceptible lines, 80 mg/kg) repeated over 3–5 successive generations of the parasite (Cioli and Pica-Mattoccia, 1984; Dias and Olivier, 1985). In these lines, resistance was heritable and stable even in the absence of drug pressure over 9–30 generations (Dias and Olivier, 1985; Cioli et al., 1989; Drescher et al., 1993). Classic genetic crosses between sensitive and resistant strains, as well as genetic complementation demonstrated that a single recessive autosomal locus codes for the hycanthone/oxamniquine resistance in two strains, MAP (named from the initials of the patient from which it was isolated) and Baltimore Rome Resistant (BRR) (Cioli and Pica-Mattoccia, 1984; Cioli et al., 1989, 1992; Pica-Mattoccia et al., 1992b, 1993). Similar results were also obtained *in vitro* by Coles and Bruce (1987). They cultured schistosomulas (an immature developmental stage occurring once cercariae have entered the mammalian host) under 240 p.p.m. of oxamniquine, a lethal concentration for 99.8% of the population. Survivors were injected in mice, and their progeny was not responsive to very high concentrations (500 mg/kg) of the drug. Experiments with tritiated hycanthone and oxamniquine demonstrated that the molecules were absorbed by all worms, but did not persist in resistant worms, while they were found to be covalently bound to DNA, RNA, and proteins in the sensitive strains (Pica-Mattoccia et al., 1988, 1989). Biochemical characterization showed that an enzymatic activation, probably by a sulfotransferase (Pica-Mattoccia et al., 2006), was needed by both oxamniquine and hycanthone to be lethal to the worm. Prior to its formal characterization, Pica-Mattoccia et al. (1992a) were able to biochemically isolate a fraction containing this enzyme from ground sensitive worms. They showed that resistant worms were killed when exposed *in vitro* to either hycanthone or oxamniquine and to the sulfotransferase-containing fraction from sensitive worms, while unaffected when exposed to one of the drugs alone. In the same article, the authors established that adult male worms, that are

more sensitive to the drugs than females, have a higher enzymatic activity than females and immature stages (mostly unaffected by the drugs at the classical therapeutic concentration, and having a very low enzymatic activity). A difference in the enzyme specificity might also explain why hycanthone, but not oxamniquine, is efficient against *Schistosoma haematobium* (Pica-Mattoccia et al., 1997). The sequencing of the genome of *S. mansoni* (Berriman et al., 2009) provided new opportunities of investigation, and Valentim et al. (2013) recently used linkage mapping methods to identify the gene responsible for resistance in the HR and MAP strains. Candidate genes were screened using recombinant proteins, RNAi, and protein crystallography. The authors showed that it was indeed a sulfotransferase that was responsible for the resistance. They identified its coding gene (Smp_089320) and the non-synonymous nucleotide substitutions causing the loss of function of the enzyme in resistant lines MAP and HR, thus clearly establishing the mechanism of resistance in the studied lines (see Figure S1 in Valentim et al., 2013 for the mode of activation of oxamniquine).

Nevertheless, there are several observations that suggested early that alternative mechanisms might exist. For instance, Jansma et al. (1977) described three different ways to induce resistance by collecting the progeny of worms after a single injection of 60 mg/kg of hycanthone to the rodent host 54–70 days post infection (Type I), or 27–29 days post infection when worms are still at an immature stage (Type II), and from hosts infected by a single sex and then re-infected 2–58 weeks later by cercariae of the opposite sex (Type III). In all three cases, parents were all susceptible to the drug while a high percentage of the progeny survived the administration of a curative dose of hycanthone. Resistance was shown to be metastable to various degrees (depending on the type of induction used) through 10–21 generations. By metastable, we mean that from one generation to another, the percentage of worms surviving hycanthone treatment was highly fluctuant and did not follow a classical Mendelian inheritance pattern. For example, it was observed at several occasions by Jansma et al. (1977) that while a parental generation was completely resistant (100% survivors), a high proportion of the direct progeny (sometimes up to 75%) was found sensitive. Sometimes, the amount of resistant parasites would go up again a couple of generations later. These results are in complete contrast of the work mentioned above where classical recessive monogenic pattern for the inheritance of the resistance was observed. The work of Jansma was partly reproduced by Brindley et al. (1989). They were able to achieve Type II induction with lines of schistosomes deriving from Jansma's isolate.

As these observations do not fit the monogenic resistance model recently described, we reasoned that environmentally triggered drug resistance with metastable inheritance is a legitimate candidate for an epigenetically based phenotype. We use here a more narrow definition of epigenetics, meaning heritable changes in gene expression, without changes in the DNA sequence (Wu and Morris, 2001). They can be physically based on chemical modification of the DNA (e.g., methylation of cytosines), changes of proteins interacting with DNA (e.g., histone modifications), as well as nuclear localization of chromosomes, and short untranslated RNA playing a role in post-transcriptional silencing of

genes (Aguilera et al., 2010). There are numerous examples in the literature on how environmental factors, such as nutrition, pollution, temperature, and others, can have an influence on epigenetic control (Feil and Fraga, 2012). A drug treatment could, for example, induce an epimutation, silencing the gene coding for the enzyme responsible of the bio-activation. This epimutation would be more or less stable and/or inherited to progeny given the parasite strain (as the phenotype was previously described to be either stable or metastable through generations depending on investigators and strains). This type of transient transgenerational modification was initially also called “Dauermodifikation” (Jollos, 1921).

From this hypothesis, we decided to revisit the historical data and to complement it with our own study. DNA methylation has been found to regulate oviposition (Geyer et al., 2011). However, it is globally at a very low level through the genome (Raddatz et al., 2013), and therefore we decided to focus on another epigenetic information carrier, histone modifications. We decided to use a Brazilian (SmBre) and Guadeloupian (SmGH2) strain. Both strains were previously pool-sequenced (Clément et al., 2013) and we found that they were both of low genetic diversity (Tajima's π of 2.10^{-4} for SmGH2 and $1.8.10^{-4}$ for SmBRE) and do not have any coding polymorphism in the gene (Smp_089320) uncovered by Valentim et al. (2013). Even if there is an almost uniform genetic background within the strains, we could not, at this stage completely rule out that an induction phenotype could come from the selection of a novel resistance allele in some schistosomes and not an epigenetic change. The debate whether heritable characters are induced by the environment or whether they occur randomly and are subsequently selected is not new. In 1951, in order to differentiate between “spontaneous (genetic) mutation and natural selection” and “directed mutation” or induction, the Lederbergs had designed their now famous experiment using replica plating of bacterial colonies from non-selective support to Petri dishes with selective media (Lederberg and Lederberg, 1952). Their work gave results in favor of the spontaneous mutation, i.e., a genetic variability (diversity) based mechanism. We reasoned that it would be in principle possible to reproduce the classical Lederberg experiment, but this time using clonal populations of *S. mansoni* instead of bacterial clonal colonies.

Our results show that one of the strains, SmBre, is resistant to hycanthone in the absence of mutations in Smp_089320. After exposure to the drug, the chromatin structure is altered around genes belonging to gene ontology (GO) groups involved in catalytic activity and response to stress. This finding suggests that epigenetic mechanisms could be responsible for adjustments of the parasites metabolism leading to temporary resistance, thus reconciling the seeming contradictions between stable and metastable types of the phenotype.

MATERIALS AND METHODS

MAINTENANCE OF THE PARASITES

The *Schistosoma mansoni* strains SmBre and SmGH2 were originally sampled in Brazil and Guadeloupe, respectively. Both strains were maintained in their sympatric intermediate host strain BgBre and BgGua of the mollusk *Biomphalaria glabrata* and *Mus musculus* as definitive vertebrate host. SmBre was originally

sampled in the 1960s in Recife, and SmGH2 was isolated on Nov. 10th, 1983 from a patient in Ste. Rose (Guadeloupe, French West Indies) (Pers. Commun. A. Théron).

PRELIMINARY DRUG SCREENING OF THE STRAINS

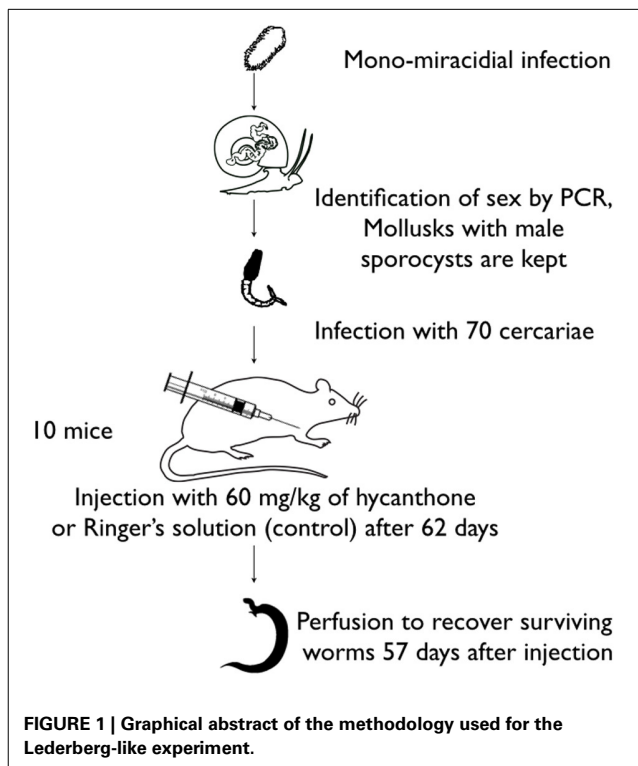
Both SmBre and SmGua were screened for resistant individuals by first infecting five mice (for each strain) with 70 mixed-sex cercariae. Mollusks from which cercariae were collected were initially infected with 20 miracidia each, and hence cercariae shed are of multiple genotypes. Mice were injected 62 days later with a dose of 60 mg/kg of hycanthone (similar to the one used by Jansma et al., 1977). Eight weeks later, the mice were sacrificed by a lethal intraperitoneal injection of sodium pentobarbital, and surviving adult worms were recovered by retrograde perfusions of the hepatic portal system with citrate (7.5%) saline (8.5%) solution administrated through the left ventricle (Duvall and DeWitt, 1967). Worms trapped in the liver or mesenteric system were collected after excising these organs.

LEDERBERG-LIKE EXPERIMENT

Biomphalaria glabrata BgBre individuals were infected with single miracidium from SmBre. The experiment was repeated with several different *B. glabrata* BgBre host individuals, each one infected with a different *S. mansoni* clone (each miracidium came from different parents). We genotyped with sex markers cercariae shed by each *B. glabrata* following Beltran et al. (2008), and we selected five mollusks infected with male parasites. We chose to work only on male schistosomes as it has been shown that they are more sensitive to hycanthone (Pica-Mattoccia et al., 1992a) than females, and to avoid sex-based bias in the epigenome analysis. From each snail, we infected 10 mice with 70 cercariae each. Sixty-two days later, half of the mice were injected with a dose of 60 mg/kg of hycanthone, and the other half with Ringer's solution as a negative control. Fifty-seven days after injection, we perfused the mice following the method of Duvall and DeWitt (1967) to collect control and surviving, resistant worms. A graphical abstract of the methodology is on Figure 1.

CHROMATIN IMMUNOPRECIPITATION AND SEQUENCING

Native ChIP was performed following the protocol developed for *Schistosoma mansoni* by Cosseau and Grunau (2011) (also available online at http://methdb.univ-perp.fr/cgrunau/methods/native_chip_sm.html) on a pool of six resistant and a pool of 10 control adult worms. Immunoprecipitation was performed using the following antibodies: H3K4me3 (Millipore, cat# 04-745 lot# NG1680351, 4 μ l per reaction), H3K9Ac (Millipore, cat# 07-352 lot# DAM16924924, 8 μ l per reaction), H3K9me3 (Abcam, cat# ab8898 lot# 733951, 4 μ l per reaction), and H3K27me3 (Diagenode, cat# pAb-069-050 lot# A29900242, 8 μ l per reaction). ChIP products were sequenced as single-end 50 bp reads on an Illumina HiSeq 2000 at the Montpellier GenomiX facility (<http://mgx.cnrs.fr>). We performed PCR amplification of the two exons of Smp_089320 on control and resistant ChIP products to test if a possible somatic mutation had occurred in that gene. We used Promega GoTaq Flex (cat# M8291) at 3 mM MgCl₂ and with the following PCR conditions: initial denaturation 2 min at 95°C followed by 35 cycles of denaturation 95°C



for 30 s, annealing 62°C for 30 s and elongation 72°C for 1 min. A final extension of 72°C for 5 min completed the reaction. We used two primer pairs to amplify exon 1 (Fwd1 tccacctctctcactcaatg, Rev1 ccacatgttgtaaatccgta, Fwd2 gctggtctaccgagaactgg and Rev2 tcccgtagaaaaccaactcg) and one pair for exon 2 (Fwd agtc-cattcattcaatgtttcaa and Rev caatccacaaatccccattc). PCR products were gel purified using Promega Wizard SV Gel and PCR Clean-Up System (cat# A9282) and sequenced at the Genoscreen facility (<http://www.genoscreen.fr/>).

QUALITY CONTROL AND ALIGNEMENT

All data treatment was carried out under a local galaxy instance (Goecks et al., 2010) (<http://galaxy.univ-perp.fr>). We used the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) for verification of the reads quality. Read quality was judged sufficiently good (the majority of reads showed a fastq-sanger quality score ≥ 24 for all positions). For the analysis of unique sequences, reads were aligned on the *Schistosoma mansoni* genome assembly v5.0 (Protasio et al., 2012) with Bowtie2 (Langmead and Salzberg, 2012) using parameters—sensitive -k 2. Reads with unique match were filtered with samtools (Li et al., 2009) (samtools view -hS -q 255). For peak calling, an equal number of randomly selected reads was chosen between the treated and control conditions (H3K4me3 = 2.9 million, H3K9Ac = 23 million, H3K9me3 = 21 million, H3K27me3 = 17.9 million). Peak identification was performed using PeakRanger v1.16 (Feng et al., 2011) with *P*-value cut off 0.0001, FDR cut off 0.01, Read extension length 200, Smoothing bandwidth 99 and Delta 1. Wiggle files generated by PeakRanger were uploaded in our local instance of GBrowse

(Stein et al., 2002) (<http://genome.univ-perp.fr>) for visualization. For the analysis of repetitive sequences, reads were aligned with Bowtie2 evoking parameters—sensitive -k 5. The 3,145 repetitive consensus sequences used as reference in this study were obtained from a previous analysis done with the same strain SmBre (Lepesant et al., 2012). From the resulting SAM file, we choose the position with the highest score for each read. Normalization of SAM files was done by randomly removing a given portion of the counts to achieve the same effective library size (H3K4me3 = 0.7 million, H3K9Ac = 9.8 million, H3K9me3 = 10.8 million, H3K27me3 = 13.9 million). Elements that showed differences in histone modification levels were annotated using the ABBlast search engine within the RepeatMasker program (Smit et al., 2010). The search was performed for both nucleotide and protein databases. Sequences without match were compared against the nucleotide database of NCBI using the blastn search engine and against the non-redundant protein database using blastx (Altschul et al., 1990).

COMPARATIVE ANALYSIS

To find differences in the histone mark pattern between control and treated conditions, we used the Pycoenrich function (with -binsize 0.3) from Pyicos v2.0.6 (Althammer et al., 2011). Visual inspection was performed on all the regions detected by the software with a *z*-score above 7. The GO enrichment analysis was performed using BLAST2GO (Conesa et al., 2005). For the analysis of significant differences in histone isoform enrichment in repeats, normalized read counts were compared using the DESeq package of the R software (Anders and Huber, 2010). This package uses the negative binomial distribution to calculate the *P*-value for each element between two compared samples and find elements that present significant differences in their enrichment levels.

GO ENRICHMENT ANALYSIS

The 60 genes for which chromatin structure differences had been detected were annotated using Blast2GO (Conesa et al., 2005). In order to check if GO terms would be overrepresented in this group of genes, 60 transcripts were randomly sampled out of 41,669 transcripts generated by a combination of RNA-Seq data and Sanger annotation. RNA-Seq data were generated from several biological replicates and these data will become available when the analysis is fully finished. The random samples were annotated as above and GO enrichment analysis was performed using Fisher's exact test implemented in Blast2GO and a *p*-value of 0.2 as cutoff. Random sampling was repeated 3 times.

RESULTS

A STRAIN WITH SENSITIVE GENOTYPE DISPLAYS HYCANTHONE RESISTANCE

We compared the number of worms recovered in mice injected with hycanthone to those recovered from control mice. We found 25 (36%) surviving worms, perfused from 5 mice, in the Brazilian strain SmBre, despite the fact that it has the sensitive genotype for Smp_089320. The Guadeloupian strain SmGH2, however, is completely sensitive to hycanthone, i.e., no survivors were found in any of the five infected and treated mice.

CLONAL SCHISTOSOMES RESPOND DIFFERENTLY TO HYCANTHONE TREATMENT

Following the discovery of hycanthone resistant individuals within the SmBre population, we adapted the Lederbergs' experiment to our schistosome model. To do so, we infected five *Biomphalaria glabrata* with a single, genetically distinct, SmBre miracidia. Cercariae shed by each mollusk are genetically identical. We then infected 10 mice with 70 cercariae. Half of the mice were then treated with an injection of hycanthone similar to the one described by Jansma for Type I induction (Jansma et al., 1977), and the other half was injected with Ringer's solution, as negative controls. As we excluded the possibility of a somatic mutation in Smp_089320, similar to the classical replica plating experiment, two possible outcomes can be imagined: (1) if a random mutation had occurred in one of the clones that conferred resistance to hycanthone (pre-existence), then only cercariae from this very snail would produce resistant flukes because they are genetically identical (**Figure 2A**) (2) if the phenotype is induced through the treatment with a certain probability, we would observe resistance randomly associated with different mice hosts and without correlation with the particular snail host (similar to Lederbergs' bacterial colony, **Figure 2B**). We later perfused the mice to look for control and surviving, resistant worms. The results of our experiment are shown in **Figure 3**. In summary, while the vast majority of mice were exempt of parasites after hycanthone injection, we found surviving worms in six different rodents, belonging to three different *S. mansoni* genotypes (B–D). We did not find any worm originating from genotype A and E after the treatment. In contrast, under the control condition, we recovered an average of 30 ± 5 worms (confidence interval at α 5%) per mice for the five genotypes. Sequencing of Smp_089320 in resistant worms did not show any of the three mutations identified by Valentim et al. (2013) or any new polymorphism.

CHROMATIN STRUCTURE CHANGES ASSOCIATED WITH THE RESISTANT PHENOTYPE

We performed ChIP-Seq on the pool of the six hycanthone resistant worms, as well as on control worms. We targeted

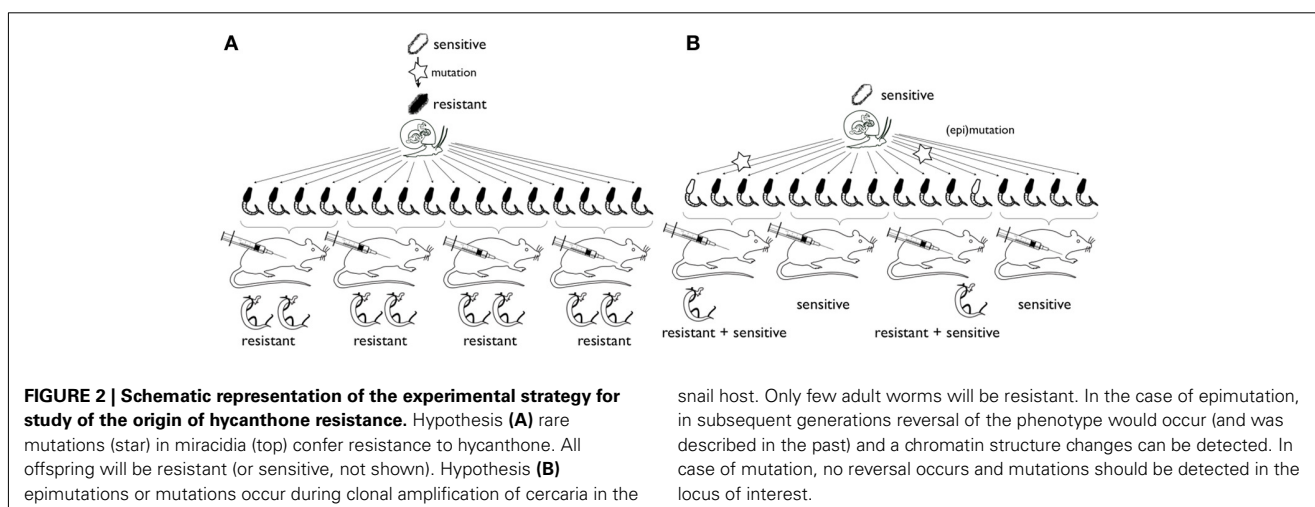
four modifications of histone H3: tri-methylation of lysine 4 [H3K4me3], tri-acetylation of lysine 9 [H3K9ac], tri-methylation of lysine 9 [H3K9me3], and tri-methylation of lysine 27 [H3K27me3]. The first two are associated with euchromatin (relaxed DNA structure allowing transcription) and the later two are associated with heterochromatin (condensed DNA structure impeding transcription). Around Smp_089320 none of the marks showed enrichment. We then searched for variations in the chromatin structure at the genome level, and a bioinformatic analysis, followed by visual inspection, was able to detect differences in the histone modification profiles between hycanthone resistant and control worms. We found 47 differences for H3K4me3, 43 of which being located at the transcription start site of genes. This corresponds roughly to 0.6% of the total peaks detected for this histone mark. For H3K9me3, 12 differences were spotted (0.012% of total peaks), mostly in introns and within 2 kb upstream and downstream of genes. We identified five differences for each of the other two marks (0.007% and 0.005% of total peaks for H3K9Ac and H3K27me3, respectively), the majority being outside of genes. A list of the regions and associated genes in which we found differences is in Supplementary Table 1.

CHROMATIN STRUCTURE MODIFICATIONS ARE PREFERENTIALLY ASSOCIATED WITH GENES RELATED TO CATALYTIC ACTIVITY

None of the genes in which differences were uncovered code for a sulfotransferase. Using a GO enrichment analysis, we did not find any clear pathway linking together the genes bearing chromatin structure changes. However, we found that the terms “transferase activity” (13 occurrences), “kinase activity” (8 occurrences) and “catalytic activity” (26 occurrences) and “response to stress” (5 occurrences) were overrepresented in the panel of genes with epimutations.

CHROMATIN STRUCTURE MODIFICATIONS IN REPETITIVE SEQUENCES

We identified, in 14 repetitive sequences, statistically significant changes in chromatin structure. This is very few (0.44% of repeats) showing that globally the chromatin remains stable. Interestingly, changes in H3K4me3 and H3K9ac always decrease



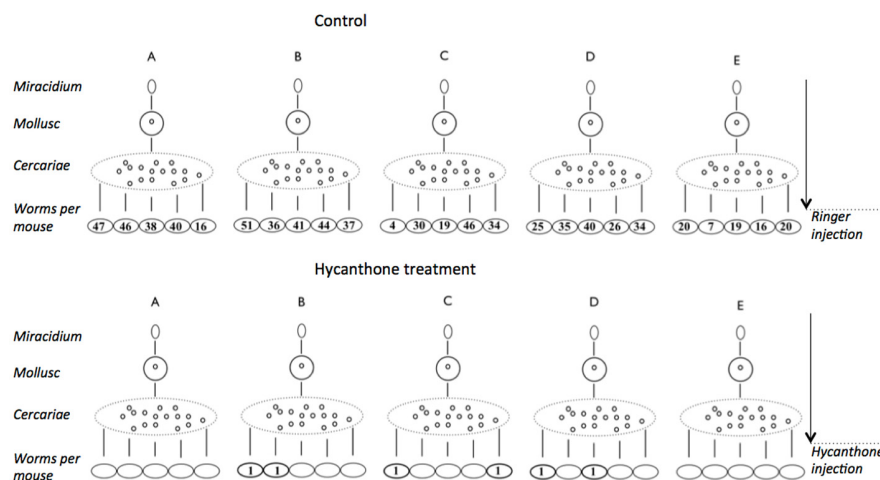


FIGURE 3 | Results of the experiment on induction of hycanthone resistance using mono-miracidial infections. Five mollusks were infected using a single miracidium (genotypes A–E). Cercariae shed from a single mollusc are clonal and were used to infect 10 mice. Half of the mice were injected with a control Ringer's solution and

the other half with a therapeutic dose of hycanthone. Mice were later perfused and worms were counted. In the hycanthone condition, six worms of three genotypes (B–D) were recovered from six mice, while an average of 30 worms per mouse were found in the control condition.

in the treated worms while H3K27me3 increases, indicating a tendency toward heterochromatinization. No changes occur in H3K9me3. H3K4me/H3K9ac modifications are observed in dispersed repeats that belong to the LINE/LTR classes. In contrast, H3K27me3 becomes enriched in simple repeats that are organized in blocks (Supplementary Table 2) sometimes covering the entire length of large contigs for which chromosomal location is unknown.

DISCUSSION

When screening the literature on hycanthone resistance, we were surprised to see such divergent results between the different research groups. On one hand, there is clearly established monogenic, stably inherited, resistance, and on the other, there is an inducible resistance, appearing from sensitive parents, and not inherited in a Mendelian fashion. We took a deeper look at the methodology employed by the all the research groups, and their interpretation of data, with the hope of finding variations in the procedures, or in the analysis methods, that could, at least partly, explain the discrepancies between the observations and conclusions.

We first tried to track back the origins of the strains that were employed in the various studies on hycanthone/oxamniquine resistance (Table 1). Our conclusion is that very few strains were used in these studies, and most of them had been maintained in laboratory for a very long time, probably reducing the genetic diversity of the strains. We identified six resistant strains in the literature, and four of them derived from the M strain used in Jansma et al. (1977) for the first induction experiments. We also found four main sensitive strains. At least four other strains (either resistant or sensitive) were punctually used and will not be discussed here (Bruce et al., 1987; Drescher et al., 1993; Pica-Mattoccia et al., 1993).

In his Type I induction, Jansma (Jansma et al., 1977) injected a dose of 60 mg/kg of hycanthone 54–70 days post-cercarial infection. He noted a marked reduction in the number of worms, but rarely a complete elimination. Surviving worms were unable to produce viable eggs for 4–8 months. After that period, eggs from which miracidia were hatched were collected to produce a F1. This first generation was treated with a therapeutic dose of hycanthone (80 mg/kg) and 94% of the worms were found to be resistant. Twenty-one successive generations were produced from this F1, and on several occasions, additional groups of mice were infected from cercariae batches shed on different days, by the same infected mollusks, leading to parallel generations known as branches. Resistance tends to decrease with generations in the main branch, and is highly variable between branches (some are totally resistant for several generations, some become completely sensitive, and others are fluctuant). Although it is not explicitly stated in the paper, it seems that progeny for generations F2–F21 was collected before the therapeutic drug injection used to estimate the percentage of resistant parent worms. Hence, there was no drug selective pressure from one generation to another, and that could partly explain the metastability observed by Jansma but not by other investigators, as suggested by Cioli and Pica-Mattoccia (1984). It is conceivable that Type I resistance was achieved not by induction, but through the selection worms with a resistant genotype, and as the induction dose is slightly below the therapeutic one (normally killing >90% of the parasites), there is possibility that some worms with a normally “sensitive” phenotype, escaped the drug. Cioli et al. (1992) observed that their resistant strain derived from Jansma's had a lower global infectivity (to mollusk and mammalian host) and a lower fertility (lower amount of eggs per females) than their sensitive strain (although this could be strain related, as the sensitive strain from Cioli is not from the same origin as Jansma's). In the hypothesis

Table 1 | Origin and resistance phenotypes of the most commonly used strains in the literature.

Strain	Origin	Resistance	Derived from?	Comment	References
M	Puerto Rico	Oxa/Hyc	–	Used for the three types of induction experiments	Jansma et al., 1977
BRR	Puerto Rico	Oxa/Hyc	M	Underwent drug selective pressure over 3 generations	Cioli et al., 1992; Pica-Mattoccia et al., 1992a,b, 1993, 2006
JHU	Puerto Rico	Oxa/Hyc	M	Successful Type II induction	Brindley and Sher, 1987; Brindley et al., 1989, 1991
H-30	Puerto Rico	Oxa/Hyc	M		Souza et al., 1988
HR	Puerto Rico	Oxa/Hyc	M		Valentim et al., 2013
MAP	Brazil, Minas Gerais	Oxa Partial to Hyc	–		Dias and Olivier, 1985; Bruce et al., 1987; Souza et al., 1988; Pica-Mattoccia et al., 1992a,b, 1993; Drescher et al., 1993; Valentim et al., 2013
SEN	Puerto Rico	None	–		Cioli, 1976
BH	Brazil, Minas Gerais	None	–	Isolated from an untreated patient in 1967	Dias and Olivier, 1986; Bruce et al., 1987
NMRI	Puerto Rico	None	–	Isolated in the 1940'	Fletcher et al., 1981; Brindley and Sher, 1987; Brindley et al., 1989
LE	Brazil	None	–		Valentim et al., 2013

Hyc, Hycanthone resistant; Oxa, Oxamniquine resistant.

that some “sensitive” worms survived the induction step and were mixed with selected resistant worms, this could explain why the resistance is does not follow a Mendelian inheritance. This cannot be the case in Type II, as immature worms have little or no sensitivity to hycanthone (induction was also successful with sub-curative doses ranging from 3 to 60 mg/kg) and in Type III where the parents were not exposed to the drug before collecting the progeny. These observations lead us to think that the resistant phenotype in Type II and III could indeed be induced and transmitted to the progeny to a certain extent. Cioli and Pica-Mattoccia (1984), Cioli et al. (1989) and Dias and Olivier (1986) stated that they were unable to reproduce Jansma's results in inducing Type II or III resistance. Three other antihelminthics, oxamniquine, praziquantel, and oltipraz were also unsuccessfully tested with the hope of inducing Type II resistance and selecting a new line of resistant schistosomes (Dias and Olivier, 1986). However, we have a slightly different interpretation of the results from Dias and Oliver. In their induction protocol, mice were treated with subcurative dose of one of the three drugs, 26–30 days after cercarial infection (when schistosomula have a low sensitivity to both hycanthone and oxamniquine), and then divided into two groups. The first group was treated with a curative dose of drug, while the second was left as a control, and the eggs from that control group were used to restart a cycle of the parasite. Mice infected by the F1 cercariae from the control group were also divided into a trial group treated with a curative dose of the drug and a control group (which would serve to generate a F2). No resistant worms were found in the treated F1 and F2, but 8–12% were found in the parental generation (in which the induction step was performed). Admittedly, there was no inheritance of a resistance phenotype to F1 and F2, but the fact that survivors were found only in the parental generation which underwent the induction

step let us believe that it was not as a failure at inducing resistance as described by the authors. In fact, only when induction was performed, some resistant worms were found. Type II was also successfully induced by the Brindley group (Brindley et al., 1989) in several lines of the parasite deriving from Jansma's isolate. However, while four strains tested in this article were described as sensitive by the authors, it is worth to mention that in some of the trials, up to 26% of the worms survived the hycanthone treatment (with or without preliminary Type II induction). Brindley also observed spontaneous occurrence of resistance in other strains with no induction, although he admitted that “*whether these fluctuations reflect true changes in the parasite genotype rather than experimental variation was not systematically investigated in this study*” (Brindley et al., 1989). This lead us to realize that for the majority of research groups, resistance was considered as achieved when at least 90% of the worm population survived a therapeutic dose. In several studies (Dias and Olivier, 1986; Brindley and Sher, 1987; Brindley et al., 1989; Drescher et al., 1993), various lines of schistosomes did not reach this resistance threshold (with or without induction and/or drug pressure over a couple of generations). They were considered as sensitive although in some cases there was a non-negligible percentage (3–26%) of worms surviving a therapeutic dose of either hycanthone or oxamniquine, strengthening our hypothesis that a mechanism following a non-Mendelian inheritance pattern is involved. In this context, we believe it is more appropriate to talk about the penetrance of the phenotype rather than talking strictly about sensitivity or resistance.

Also, Cioli and Pica-Mattoccia specifically mentioned that their approach was different from the Jansma/Bueding/Brindley group. Three generations of selection were necessary to obtain a phenotype that was stably inherited by the progeny (Cioli and

Pica-Mattoccia, 1984). The same observation was made by Dias and Olivier (1985), as in their experiments, resistance was stable only after the 5th generation (experiments were continued until 14th generation). Unfortunately, the data for the first 3–5 generations are not available but one can assume that incomplete penetrance of the phenotype was observed. These large differences in penetrance of resistance phenotypes (i.e., large dose dependency) in different isolates were also found by others (Coles et al., 1986; Kinoti, 1987) and also described in Pica-Mattoccia et al. (1993). Kinoti noted important variations in the dose of oxamniquine needed to eliminate 99% of the worms (ED99), in both laboratory and clinical trials. For example, East African isolates needed a dose of oxamniquine 200–250% higher than Puerto-Rican isolates to reach ED99 (Kinoti, 1987). In all publications, at least 3 generations of selection were necessary to produce stable resistance, but even then, phenotypic differences between the different resistance strains and most notably an exception to the Mendelian segregation pattern were observed. In the MAP strain for instance there was (1) a difference between heritability of hycanthone and oxamniquine resistance and (2) segregation did not follow the recessive pattern. Segregation could not be followed up because the females produced no eggs, i.e., no genetic analysis was possible (*“Thus, we are left with no obvious explanation for this unexpected partial exception to the overall recessivity of the MAP resistance.”*) (Pica-Mattoccia et al., 1993). The reason for the incomplete penetrance in the first 3–5 generations cannot be determined with confidence from the published data alone. It might be that the initial strains showed a high degree of genetic diversity with only a few individuals carrying the mutation in the sulfotransferase gene. Molecular biology methods in *S. mansoni* research were not sufficiently advanced in that time to measure genetic diversity. From the few available data, however, it seems that the founder populations of the resistant strains were very small. Years of maintenance cycle in laboratories, as well as the possibility of a founder effect every time that the strains were transferred or shared with other laboratories let us think that the genetic diversity was probably low in the different strains. We also noted that most of the strains were from the same geographical areas (either Puerto Rico or the region of Minas Gerais in Brazil) and represent only a small share of geographic diversity of *S. mansoni*. Another reason for variability in penetrance could therefore be high phenotypic plasticity and/or variability in plasticity. Jansma's results were in favor of this latter mechanism, and he additionally described transgenerational plastic effects. It has also to be noted that while the work of Brindley and Jansma was performed using hycanthone, other research groups used oxamniquine. Although schistosomes were always cross resistant or cross sensitive to both molecules, hycanthone is known to have a mutagenic effect, and it could have played a role in the phenomenon of induction, which was never observed with oxamniquine.

In our own experiments, we worked with a strain bearing the sensitive genotype for the sulfotransferase gene previously identified as the source of hycanthone/oxamniquine resistance, but we still found in drug screening that in one strain (SmBre) about a third of the adult worms survived the treatment while another strain did not show survivors. Resistance of SmBre

could be due to a mutation in a different locus and/or due to epimutations but this would still not explain Jansma's induction phenomenon. Therefore, we used a Lederbergs experiment to discriminate if the resistance was pre-existing or induced by the experimental conditions. Similar to the classical replica plating experiment, two possible outcomes could be imagined: (1) if the phenotype is induced through the treatment with a certain probability, we would observe resistance randomly associated with different mice hosts and without correlation with the particular snail host (similar to Lederbergs' bacterial colony) (2) if a random mutation had occurred in one of the clones that conferred resistance to hycanthone (pre-existence), then only cercariae from this very snail would produce resistant flukes because they are genetically identical. A caveat with the Lederberg experiment in organisms with epigenetic inheritance systems is the potentially heritable transcriptional and/or posttranscriptional level of diversity. It is possible that epimutations occur during clonal amplification of the cercariae (**Figure 2B**). Also, as long as somatic mutations cannot be excluded, mutation would do the same job. Both would mimic the outcome of the “induction” alternative. We found that not all the worms issued from the same miracidia (i.e., being genetic clones of each other), responded to the treatment the same way. While the vast majority was sensitive, we still found some resistant individuals. When interpreted in the context of Lederberg's experiment, would favor the directed (epi)mutation hypothesis. If we had selection of a pre-existing mutation or resistance phenotype, then we would have expected that all, or no, worms from a same genotype would present the same phenotype (resistance or sensitivity). From our results, we propose the hypothesis that epigenetics could be part of the mechanism underlying the resistance induction.

Our ChIP-Seq experiment provided evidences of epigenetic variation between resistant and control worms. Our analysis highlighted at least 64 chromatin structure changes between the two conditions, with an enrichment profiles in genes implicated in catabolic, detoxification, and transfer activity (Two examples can be seen in Supplementary Figure 1). No specific pathway linking these genes together arose, but it makes sense in the context of a response and the subsequent elimination to a drug to see these kinds of biological processes being involved (as stated before, hycanthone is known to be eliminated from resistant worms a certain amount of time after its penetration, while it is stuck in sensitive worms metabolism). It is therefore conceivable that the drug resistance observed in the surviving worms is due to epistatic interaction of multiple gene products rather than an (epi)mutation in a single locus. Another interesting finding is the presence of a differential H3K27me3 profile in the gene body of SmMRP1 (Multidrug Resistance Associated Protein 1, Smp_171740). In various species, including trematoda, this gene is part of a family of transporters playing a role in exclusion and elimination of xenobiotic compounds and metabolic toxins (Kasinathan et al., 2010). In schistosomes, it was demonstrated that SmMRP1 has higher levels of expressions in response to the widely used schistosomicide praziquantel, and possibly other drugs (Kasinathan et al., 2010; Greenberg, 2013). We found a significant enrichment in H3K27me3 around exon 21 of this

SmMRP1 in the control condition. Enrichment of this mark are usually associated with increases of facultative heterochromatin and gene repression (Trojer and Reinberg, 2007; Bannister and Kouzarides, 2011). Our current knowledge of the functional role of histone modifications in gene expression in *S. mansoni* does not allow us to affirm that the lower presence of this mark in the resistant worms would lead to an increased expression of this gene, but it stays a promising candidate for further experiments.

Fourty-seven percent of the genome of *S. mansoni* is composed of repeats (Lepesant et al., 2012). The function of these repeats is completely unexplored and, in general, they are not included in the analysis of molecular basis of phenotypic variation. We have shown earlier (Lepesant et al., 2012), that repeats can be of importance in schistosomes, and we decided to study their chromatin structure changes upon hycanthone treatment. Only a very small fraction of the repeats (0.4%) change their histone modification status. In contrast to other organisms, stress (in this case hycanthone treatment) does not seem to lead to an euchromatinization and potentially mobilization of repeats but has the inverse effect (decrease of H3K9ac and increase in H3K27me3). Interestingly, the repeats in which H3K27me3 status changes are located in satellite-type blocks on specific locations on the genome. It is conceivable that they form knobs that separate genome domains and influence indirectly gene expression. Further work will be needed to understand the impact of these repeated regions on the definition of the chromatin structure in which they are included.

While this ChIP-Seq experiment brings new insight on epigenetic mechanisms and regulation in *S. mansoni*, it does have some limitations. In order to be as faithful as possible to Jansma's work to reproduce his Type I induction, we exposed the worms to hycanthone through an injection in the rodent host. However, this generates a degree of variability from one mouse to another, as each individual has a different metabolism, immune system, clearance rate, and other physiological factor influencing the exposure of the worms to the drug. Also, although cercariae issued from a same initial miracidium are theoretically clonal, genomic instability such as somatic mutation and recombination can occur during asexual life stages like sporocysts, as shown by Vieira et al. (1991), and Bayne and Greveling (2003), and their impact on the phenotype should be taken into account. Nonetheless, we managed to recover worms that survived the hycanthone treatment at two occasions (initial drug screening and Lederbergs-like experiment), from a low genetic diversity strain and without any evidences that the gene uncovered by Valentim et al. (2013) was involved. This, in addition with the differences of chromatin structure observed between surviving and control worms lead us to think that induction of resistance in *S. mansoni* is not a myth after all. Nevertheless, we do not think that induction should be seen as a specific response of the epigenome to a certain environmental condition leading to over/under-expression of a specific gene. Instead, this induction phenomenon seems to come from the capacity of some strains, or some individuals within a strain, to extend their degree of plasticity beyond the genetically defined "default" state through chromatin structure changes. This allows them to overcome an environmental stress in the absence of a pre-existing stress-response phenotype. Depending on the length of the stress, we can imagine that

the progeny can more or less efficiently inherit the epigenetic condition responsible for the phenotype until it is fixed in the population (through a DNA mutation for example). This fits well with theoretical framework developed recently in which heritable phenotypic variation is based on at least two components: the low-fidelity (epigenetic) system and the high-fidelity (genetic) system (Klironomos et al., 2013). In this model, upon changes in the environment, the low-fidelity system can generate new phenotypes that can explore the adaptive landscape. The phenotype is decoupled from the genotype and appears before the adaptive genotype. The low fidelity of the epigenetic system means the adaptive phenotype could theoretically be lost as quickly as it appears, unless a strong selective pressure is applied. This is consistent with the observations of the metastability of hycanthone resistance in absence of drug pressure (Jansma et al., 1977; Brindley et al., 1989), and the stability of the phenotype after 3–5 successive generations of schistosomes where under drug pressure (Cioli and Pica-Mattoccia, 1984; Dias and Olivier, 1985). Basically, epigenetic variation allows to "buy time" for a few generation (in this case, 3–5) until a stable, genetically transmitted phenotype occurs. To understand better the underlying mechanism, measuring the expression level of Smp_089320 and of candidate genes identified here would be informative in future studies.

Since hycanthone is not anymore used for treatment of bilharzia and the use of oxamniquine is depreciated (Galdino da Rocha Pitta et al., 2013), mechanism of resistance is more of academic interest. However, experience with hycanthone and oxamniquine shows that a new conceptual framework is necessary to design experiments that distinguish the part of genetics and epigenetics in resistance phenotypes. We believe that a standard procedure would be beneficial for the testing of resistance to current drugs, especially praziquantel, the only molecule actually used for mass treatment of schistosomiasis. Resistance to praziquantel shares striking similarities with early observations on hycanthone/oxamniquine. It can be induced through subcurative, but increasing doses of praziquantel over six-seven generations, and genotyping of two populations of schistosomes showed that there was no significant difference in the genotypes before and after drug treatment, meaning no selection of specific genotypes by the drug (Blanton et al., 2011; Greenberg, 2013). We think that a standard procedure for the next drugs should be similar to the one we used to detect resistant individuals from a clonal population. Mono-miracidial infection would be a good start, as it minimizes the genetic variation between individuals and allows to detect changes in the epigenotype. A higher number of mono-miracidial infections, using several strains of *S. mansoni* (as the induction seems to be strain dependent), and a larger number of clones (adult or schistosomula) from each mono-miracidial infection should be tested, as we still do not have a clear idea of the frequency of the phenomenon. Induction step can be done *in vivo* (for any of the tree types) but screening of resistant progeny has to be performed *in vitro*, as described by Coles and Bruce (1987) and Pica-Mattoccia et al. (1992b). This is mandatory to make sure that all parasites have the same exposure to the drug. To look for heritability, resistant worms can be surgically reinserted into new rodent hosts and allow to reproduce. It is critical to

have a standardized procedure to maintain the life cycle of the parasite (i.e., number of miracidia and cercariae, as well as their genotypes, used to infect the hosts) as it is difficult, if not impossible, in some older publications to find out if variation in the numbers of perfused worms come from effect of the drug or bias in infections. Using the same intermediate and final host is also important as they have an impact on the phenotype and life traits of the parasite. In the articles we reviewed, most experiments were done using the same final host (swiss albino mice), but sometimes hamsters were used, too. The geographical origin of the mollusk host is not always explicitly stated, and there are numerous evidences that it plays an important role in the parasite infection success and development (Theron et al., 1997). Although this was never explicitly discussed, it could be a cause to the contradictory results on hycanthone induction. Finally, both genomes and epigenomes of resistant worms would have to be compared to the ones of sensitive worms to figure out the origin of the phenomenon. Also, from our ChIP-Seq results, we saw that it is essential to keep a global, genome scale approach including repetitive sequences. The epimutations we found were distributed across the genome and are involved in various pathways. A candidate gene approach would probably have missed some of the variations we detected. Rather than only focusing on find a single gene responsible for the phenotype, we believe that the approach we describe would lead to a more complete answer on how schistosomes develop resistances.

ETHIC STATEMENT

The French Ministère de l'Agriculture et de la Pêche and French Ministère de l'Education Nationale de la Recherche et de la Technologie provided permit A 66040 to our laboratory for experiments on animals and certificate for animal experimentation (authorization 007083, decree 87–848 and 2012201-0008) for the experimenters. Housing, breeding and animal care followed the national ethical requirements.

AUTHOR CONTRIBUTIONS

David Roquis performed the literature review, achieved the bioinformatics analyses and wrote the manuscript. Julie M. J. Lepesant designed and performed the experiments, contributed to prepare the reagents and the materials. Emanuel Villafan and Cristina Vieira analyzed the data concerning repetitive DNA. Céline Cosseau conceived the experiments and took part in writing the manuscript. Jérôme Boissier contributed to drug treatments and acquisition of samples. Christoph Grunau conceived and designed the experiments, contributed to analysis tool preparation and bioinformatics analyses, and took part in writing the manuscript.

ACKNOWLEDGMENTS

We are grateful to Prof. Sam Loker (Albuquerque, NM, USA) who attracted our interest to induced hycanthone resistance and Prof. Michael J. Waring (Cambridge, UK) for sharing his precious stock of hycanthone. Without his help this work would not have been possible since the drug is not anymore commercially available. This work received support from the ANR 2010 BLAN 1720 01 (EPIGEVOL). The manuscript profited tremendously from the critical reading by two anonymous reviewers.

SUPPLEMENTARY MATERIAL

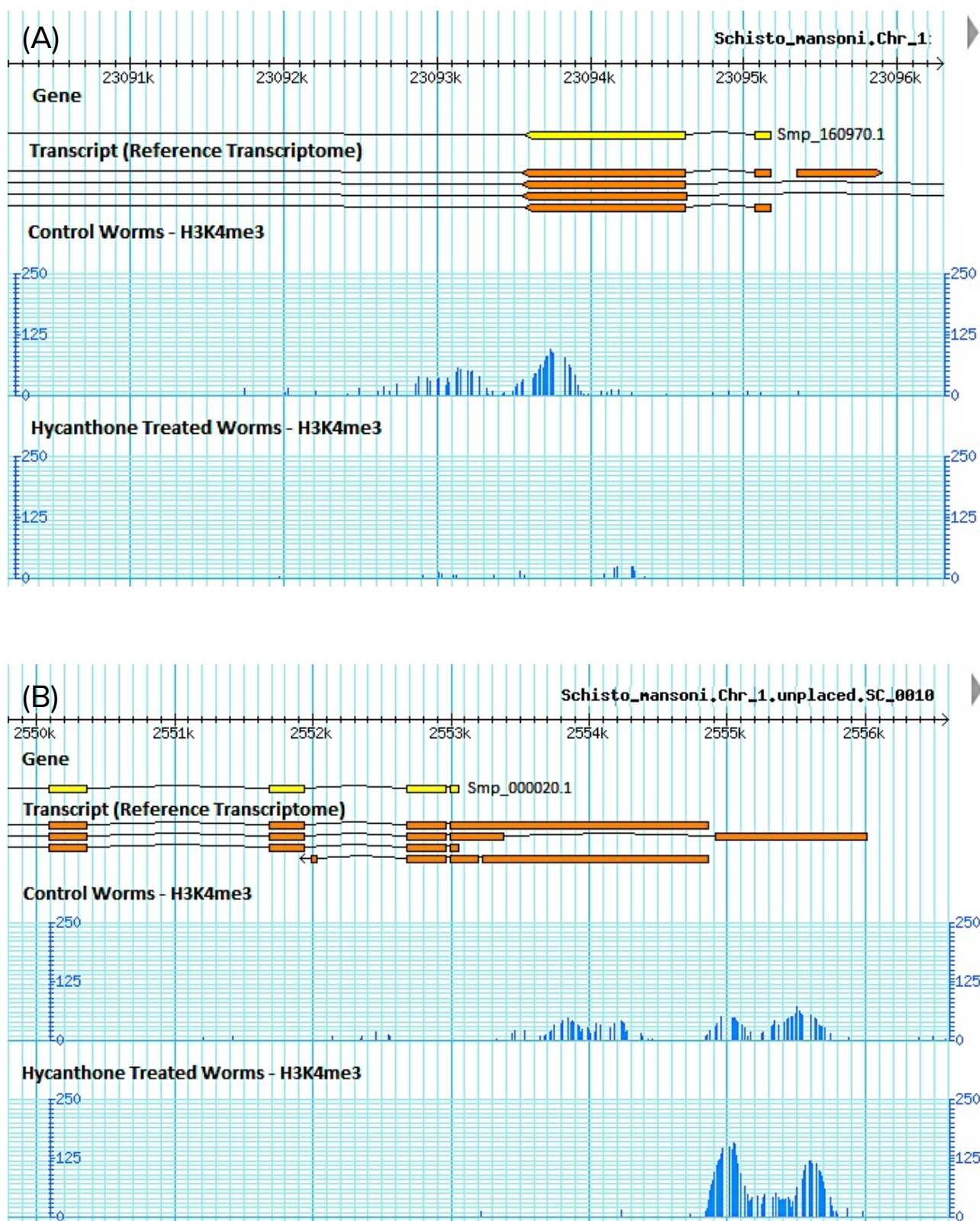
The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2014.00207/abstract>

REFERENCES

- Aguilera, O., Fernandez, A. F., Munoz, A., and Fraga, M. F. (2010). Epigenetics and environment: a complex relationship. *J. Appl. Physiol.* 109, 243–251. doi: 10.1152/jappphysiol.00068.2010
- Althammer, S., Gonzalez-Vallinas, J., Ballare, C., Beato, M., and Eyra, E. (2011). Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics* 27, 3333–3340. doi: 10.1093/bioinformatics/btr570
- Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1006/jmbi.1990.9999
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi: 10.1186/gb-2010-11-10-r106
- Bannister, A. J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395. doi: 10.1038/cr.2011.22
- Bayne, C. J., and Grevelding, C. G. (2003). Cloning of schistosoma mansoni sporocysts in vitro and detection of genetic heterogeneity among individuals within clones. *J. Parasitol.* 89, 1056–1060. doi: 10.1645/GE-3186RN
- Beltran, S., Galinier, R., Allienne, J. F., and Boissier, J. (2008). Cheap, rapid and efficient DNA extraction method to perform multilocus microsatellite genotyping on all *Schistosoma mansoni* stages. *Mem. Inst. Oswaldo Cruz* 103, 501–503. doi: 10.1590/S0074-02762008000500017
- Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P., Cerqueira, G. C., et al. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460, 352–358. doi: 10.1038/nature08160
- Blanton, R. E., Blank, W. A., Costa, J. M., Carmo, T. M., Reis, E. A., Silva, L. K., et al. (2011). *Schistosoma mansoni* population structure and persistence after praziquantel treatment in two villages of Bahia, Brazil. *Int. J. Parasitol.* 41, 1093–1099. doi: 10.1016/j.ijpara.2011.06.002
- Brindley, P. J., Heath, S., Waters, A. P., McCutchan, T. F., and Sher, A. (1991). Characterization of a programmed alteration in an 18S ribosomal gene that accompanies the experimental induction of drug resistance in *Schistosoma mansoni*. *Proc. Natl. Acad. Sci. U.S.A.* 88, 7754–7758. doi: 10.1073/pnas.88.17.7754
- Brindley, P. J., Lewis, F. A., McCutchan, T. F., Bueding, E., and Sher, A. (1989). A genomic change associated with the development of resistance to hycanthone in *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 36, 243–252. doi: 10.1016/0166-6851(89)90172-2
- Brindley, P. J., and Sher, A. (1987). Anti-schistosomal drugs: observations on the mechanism of drug resistance to hycanthone, and on the involvement of host antibodies in the mode of action of praziquantel. *Mem. Inst. Oswaldo Cruz* 82, 157–161. doi: 10.1590/S0074-02761987000800027
- Bruce, J. I., Dias, L. C., de, S., Yung-San, L., and Coles, G. C. (1987). Drug resistance in schistosomiasis: a review. *Mem. Inst. Oswaldo Cruz* 82, 143–150. doi: 10.1590/S0074-02761987000800025
- Cioli, D. (1976). Transfer of *Schistosoma mansoni* into the mesenteric veins of hamsters. *Int. J. Parasitol.* 6, 349–354. doi: 10.1016/0020-7519(76)90058-8
- Cioli, D., and Pica-Mattoccia, L. (1984). Genetic analysis of hycanthone resistance in *Schistosoma mansoni*. *Am. J. Trop. Med. Hyg.* 33, 80–88.
- Cioli, D., Pica-Mattoccia, L., and Archer, S. (1989). Resistance of schistosomes to hycanthone and oxamniquine. *Mem. Inst. Oswaldo Cruz* 84, 38–45. doi: 10.1590/S0074-02761989000500005
- Cioli, D., Pica-Mattoccia, L., and Moroni, R. (1992). *Schistosoma mansoni*: hycanthone/oxamniquine resistance is controlled by a single autosomal recessive gene. *Exp. Parasitol.* 75, 425–432. doi: 10.1016/0014-4894(92)90255-9
- Clément, J. A. J., Toulza, E., Gautier, M., Parrinello, H., Roquis, D., Boissier, J., et al. (2013). Private selective sweeps identified from next-generation pool-sequencing reveal convergent pathways under selection in two inbred *Schistosoma mansoni* strains. *PLoS Negl. Trop. Dis.* 7:e2591. doi: 10.1371/journal.pntd.0002591
- Coles, G. C. (2002). Drug resistance or tolerance in schistosomes? *Trends Parasitol.* 18, 294. doi: 10.1016/S1471-4922(02)02315-2
- Coles, G. C., and Bruce, J. I. (1987). *In vitro* selection of drug resistant *Schistosoma mansoni*. *Int. J. Parasitol.* 17, 767–771. doi: 10.1016/0020-7519(87)90057-9

- Coles, G. C., Bruce, J. I., Kinoti, G. K., Mutahi, W. T., Dias, E. P., and Katz, N. (1986). Drug resistance in schistosomiasis. *Trans. R. Soc. Trop. Med. Hyg.* 80, 347–347. doi: 10.1016/0035-9203(86)90055-6
- Coles, G. C., Mutahi, W. T., Kinoti, G. K., Bruce, J. I., and Katz, N. (1987). Tolerance of Kenyan *Schistosoma mansoni* to oxamniquine. *Trans. R. Soc. Trop. Med. Hyg.* 81, 782–785. doi: 10.1016/0035-9203(87)90032-0
- Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Cosseau, C., and Grunau, C. (2011). “Native Chromatin Immunoprecipitation,” in *Epigenetics Protocols*, ed T. O. Tollefsbol (Totowa, NJ: Humana Press), 195–212. Available online at: http://www.springerlink.com/index/10.1007/978-1-61779-316-5_15 (Accessed July 8, 2013).
- Dias, L., and Olivier, C. (1985). Stability of *Schistosoma mansoni* progeny to antischistosomal drugs. *Rev. Inst. Med. Trop. São Paulo* 27, 186–189. doi: 10.1590/S0036-46651985000400005
- Dias, L. C., de, S., and Olivier, C. E. (1986). Failure at inducing resistance to schistosomicidal drugs in a Brazilian human strain of *Schistosoma mansoni*. *Rev. Inst. Med. Trop. São Paulo* 28, 352–357. doi: 10.1590/S0036-46651986000500010
- Drescher, K. M., Rogers, E. J., Bruce, J. I., Katz, N., Dias, L. C., de, S., et al. (1993). Response of drug resistant isolates of *Schistosoma mansoni* to antischistosomal agents. *Mem. Inst. Oswaldo Cruz* 88, 89–95. doi: 10.1590/S0074-02761993000100014
- Duvall, R. H., and DeWitt, W. B. (1967). An improved perfusion technique for recovering adult schistosomes from laboratory animals. *Am. J. Trop. Med. Hyg.* 16, 483–486.
- Feil, R., and Fraga, M. F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* 13, 97–109. doi: 10.1038/nrg3142
- Feng, X., Grossman, R., and Stein, L. (2011). PeakRanger: a cloud-enabled peak caller for ChIP-seq data. *BMC Bioinformatics* 12:139. doi: 10.1186/1471-2105-12-139
- Fletcher, M., LoVerde, P. T., and Woodruff, D. S. (1981). Genetic variation in *Schistosoma mansoni*: enzyme polymorphisms in populations from Africa, Southwest Asia, South America, and the West Indies. *Am. J. Trop. Med. Hyg.* 30, 406–421.
- Galdino da Rocha Pitta, M., Galdino da Rocha Pitta, M., Jesus Barreto de Melo Rego, M., and Lins Galdino, S. (2013). The evolution of drugs on schistosoma treatment: looking to the past to improve the future. *Mini Rev. Med. Chem.* 13, 493–508. doi: 10.2174/1389557511313040003
- Geyer, K. K., Rodríguez López, C. M., Chalmers, I. W., Munshi, S. E., Truscott, M., Heald, J., et al. (2011). Cytosine methylation regulates oviposition in the pathogenic blood fluke *Schistosoma mansoni*. *Nat. Commun.* 2, 424. doi: 10.1038/ncomms1433
- Goedts, J., Nekrutenko, A., Taylor, J., and The Galaxy Team (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11:R86. doi: 10.1186/gb-2010-11-8-r86
- Greenberg, R. M. (2013). New approaches for understanding mechanisms of drug resistance in schistosomes. *Parasitology* 140, 1534–1546. doi: 10.1017/S0031182013000231
- Jansma, W. B., Rogers, S. H., Liu, C. L., and Bueding, E. (1977). Experimentally produced resistance of *Schistosoma mansoni* to hycanthone. *Am. J. Trop. Med. Hyg.* 26, 926–936.
- Jollos, V. (1921). *Experimentelle Protistenstudien: Untersuchungen über Variabilität und Vererbung bei Infusorien*. Jena: Verlag von Gustav Fischer. Available online at: <http://www.biodiversitylibrary.org/bibliography/11942#summary>
- Kasinathan, R. S., Morgan, W. M., and Greenberg, R. M. (2010). *Schistosoma mansoni* express higher levels of multidrug resistance-associated protein 1 (SmMRP1) in juvenile worms and in response to praziquantel. *Mol. Biochem. Parasitol.* 173, 25–31. doi: 10.1016/j.molbiopara.2010.05.003
- King, C. H. (2010). Parasites and poverty: the case of schistosomiasis. *Acta Trop.* 113, 95–104. doi: 10.1016/j.actatropica.2009.11.012
- Kinoti, G. K. (1987). The significance of variation in the susceptibility of *Schistosoma mansoni* to the antischistosomal drug oxamniquine. *Mem. Inst. Oswaldo Cruz* 82, 151–156. doi: 10.1590/S0074-02761987000800026
- Klironomos, F. D., Berg, J., and Collins, S. (2013). How epigenetic mutations can affect genetic evolution: model and mechanism: problems & paradigms. *BioEssays* 35, 571–578. doi: 10.1002/bies.201200169
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lederberg, J., and Lederberg, E. M. (1952). Replica plating and indirect selection of bacterial mutants. *J. Bacteriol.* 63, 399.
- Lepesant, J. M. J., Roquis, D., Emans, R., Cosseau, C., Arancibia, N., Mitta, G., et al. (2012). Combination of *de novo* assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in *Schistosoma mansoni*. *Exp. Parasitol.* 130, 470–474. doi: 10.1016/j.exppara.2012.02.010
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Pica-Mattoccia, L., Archer, S., and Cioli, D. (1992a). Hycanthone resistance in schistosomes correlates with the lack of an enzymatic activity which produces the covalent binding of hycanthone to parasite macromolecules. *Mol. Biochem. Parasitol.* 55, 167–175.
- Pica-Mattoccia, L., Dias, L. C., de, S., and Cioli, D. (1992b). Genetic complementation analysis of two independently isolated hycanthone-resistant strains of *Schistosoma mansoni*. *Mem. Inst. Oswaldo Cruz* 87, 211–214.
- Pica-Mattoccia, L., Carlini, D., Guidi, A., Cimica, V., Vigorosi, F., and Cioli, D. (2006). The schistosome enzyme that activates oxamniquine has the characteristics of a sulfotransferase. *Mem. Inst. Oswaldo Cruz* 101, 307–312. doi: 10.1590/S0074-02762006000900048
- Pica-Mattoccia, L., Cioli, D., and Archer, S. (1988). Binding of tritiated hycanthone and hycanthone N-methylcarbamate to macromolecules of drug-sensitive and drug-resistant schistosomes. *Mol. Biochem. Parasitol.* 31, 87–96. doi: 10.1016/0166-6851(88)90148-X
- Pica-Mattoccia, L., Cioli, D., and Archer, S. (1989). Binding of oxamniquine to the DNA of schistosomes. *Trans. R. Soc. Trop. Med. Hyg.* 83, 373–376. doi: 10.1016/0035-9203(89)90508-7
- Pica-Mattoccia, L., Dias, L. C. D., Moroni, R., and Cioli, D. (1993). *Schistosoma mansoni*: genetic complementation analysis shows that two independent hycanthone/oxamniquine-resistant strains are mutated in the same gene. *Exp. Parasitol.* 77, 445–449. doi: 10.1006/expr.1993.1104
- Pica-Mattoccia, L., Novi, A., and Cioli, D. (1997). Enzymatic basis for the lack of oxamniquine activity in *Schistosoma haematobium* infections. *Parasitol. Res.* 83, 687–689. doi: 10.1007/s004360050320
- Protasio, A. V., Tsai, I. J., Babbage, A., Nichol, S., Hunt, M., Aslett, M. A., et al. (2012). A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* 6:e1455. doi: 10.1371/journal.pntd.0001455
- Raddatz, G., Guzzardo, P. M., Olova, N., Fantappie, M. R., Rampp, M., Schaefer, M., et al. (2013). Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8627–8631. doi: 10.1073/pnas.1306723110
- Rogers, S. H., and Bueding, E. (1971). Hycanthone resistance: development in *Schistosoma mansoni*. *Science* 172, 1057–1058. doi: 10.1126/science.172.3987.1057
- Smit, A., Hubley, R., and Green, P. (2010). *RepeatMasker Open-3.0*. Available online at: <http://www.repeatmasker.org>
- Souza, L. C., de, D., Bruce, J. I., and Coles, G. C. (1988). Variation in response of *Schistosoma mansoni* strains to schistosomicides. *Rev. Inst. Med. Trop. São Paulo* 30, 81–85. doi: 10.1590/S0036-46651988000200005
- Stein, L. D., Mungall, C., Shu, S. Q., Caudy, M., Mangone, M., Day, A., et al. (2002). The generic genome browser: a building block for a model organism system database. *Genome Res.* 12, 1599–1610. doi: 10.1101/gr.403602
- Theron, A., Pages, J., and Rognon, A. (1997). *Schistosoma mansoni*: distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Exp. Parasitol.* 85, 1–9. doi: 10.1006/expr.1996.4106
- Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* 28, 1–13. doi: 10.1016/j.molcel.2007.09.011
- Valentim, C. L., Cioli, D., Chevalier, F. D., Cao, X., Taylor, A. B., Holloway, S. P., et al. (2013). Genetic and molecular basis of drug resistance and species-specific

- drug action in schistosome parasites. *Science* 342, 1385–1389. doi: 10.1126/science.1243106
- Vieira, L. Q., Correa-Oliveira, R., Katz, N., De Souza, C. P., Carvalho, O. S., Araujo, N., et al. (1991). Genomic variability in field populations of *Schistosoma mansoni* in Brazil as detected with a ribosomal Gene Probe. *Am. J. Trop. Med. Hyg.* 44, 69–78.
- WHO. (1998). *Report of The WHO Informal Consultation on Schistosomiasis Control*. Geneva: World Health Organization.
- Wu, C.-T., and Morris, J. R. (2001). Genes, genetics, and epigenetics: a correspondence. *Science* 293, 1103–1105. doi: 10.1126/science.293.5532.1103
- Received: 22 April 2014; accepted: 18 June 2014; published online: 16 July 2014.
- Citation: Roquis D, Lepesant MJM, Villafan E, Boissier J, Vieira C, Cosseau C and Grunau C (2014) Exposure to hycanthone alters chromatin structure around specific gene functions and specific repeats in *Schistosoma mansoni*. *Front. Genet.* 5:207. doi: 10.3389/fgene.2014.00207
- This article was submitted to *Evolutionary and Genomic Microbiology*, a section of the journal *Frontiers in Genetics*.
- Copyright © 2014 Roquis, Lepesant, Villafan, Boissier, Vieira, Cosseau and Grunau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



Supplementary Figure 1: Two examples (**A** and **B**) of chromatin structure changes between control and hycanthone treated worms on the histone 3 tri-methylated on lysine 4 (H4K4me3). Identified through the bioinformatic approach and followed by a visual inspection on our local genome browser. The gene track corresponds to the gene annotation from *Schistosoma mansoni* assembly v5.0 (Protasio *et al.*, 2012) and the transcript track refers to a reference transcriptome including data from the genome assembly v5.0 and local RNA-Seq data (unpublished).

Supplementary Table 1: List of regions and associated genes where histone marks profiles were different between hycanthone resistant and control worms. Scaffolds and positions refers to the genome assembly v5.2 of *Schistosoma mansoni*.

Histone mark	Scaffold	Start	End	Region	Associated Gene	Gene Description (From GeneDB)
H3K4me3	Schisto_mansoni.Chr_1.unplaced.SC_0135	793968	800000	Intergenic		
	Schisto_mansoni.Chr_3.unplaced.SC_0233	20698	22194	Intergenic		
	Schisto_mansoni.Chr_1.unplaced.SC_0010	2553602	2554446	TSS	Smp_000020	bifunctional protein NCOAT
	Schisto_mansoni.Chr_2	12959651	12960074	TSS	Smp_014850	NCK adaptor protein
	Schisto_mansoni.Chr_3	24461398	24461588	TSS	Smp_018990	ribosomal protein L9
	Schisto_mansoni.Chr_W	51047380	51047778	TSS	Smp_032480	cohesin subunit SA 1
	Schisto_mansoni.Chr_2	13759183	13759612	TSS	Smp_048940	Vacuolar protein sorting associated protein
	Schisto_mansoni.SC_0186	208884	209950	TSS	Smp_070110	mitochondrial 39s ribosomal protein l45
	Schisto_mansoni.SC_0150	631699	632033	TSS	Smp_070610.1	mps one binder kinase activator 1
	Schisto_mansoni.SC_0013	3568161	3568700	Intergenic		
	Schisto_mansoni.Chr_1	37187228	37187692	TSS	Smp_083740	centrin 3
	Schisto_mansoni.SC_0132	859293	859739	TSS	Smp_090030	rRNA processing protein Nop9
	Schisto_mansoni.Chr_6	4219571	4219898	TSS	Smp_091650	phospholipid transporting ATPase IIB
	Schisto_mansoni.Chr_W	44735058	44735396	TSS	Smp_096670	serine:threonine protein kinase PAK 1
	Schisto_mansoni.SC_0207	341233	342296	TSS	Smp_098370.1	uncharacterized aarF domain containing protein
	Schisto_mansoni.Chr_2	19396719	19397798	TSS	Smp_104360	hypothetical protein
	Schisto_mansoni.Chr_2	24683464	24684998	TSS	Smp_135080.1	microtubule associated protein 1A
	Schisto_mansoni.Chr_3	13232551	13233800	TSS	Smp_136410	cytoplasmic polyadenylation element binding
	Schisto_mansoni.Chr_6	16062714	16064250	TSS	Smp_136660	pro neuregulin 2
	Schisto_mansoni.Chr_3	23106561	23107580	TSS	Smp_213390	ribosomal protein s6 kinase beta 1
					Smp_213400	e3 ubiquitin-protein ligase bre1
	Schisto_mansoni.Chr_W	7385089	7385499	TSS	Smp_139090.1	DENN domain containing protein 5B
	Schisto_mansoni.Chr_W	15578849	15579931	TSS	Smp_142190	carboxypeptidase n catalytic chain
	Schisto_mansoni.Chr_4.unplaced.SC_0032	1147948	1148467	TSS	Smp_148940	sorting nexin
	Schisto_mansoni.Chr_4.unplaced.SC_0032	558558	558820	TSS	Smp_149050.1	XTP3 transactivated gene B protein
	Schisto_mansoni.SC_0214	214642	215449	TSS	Smp_151380	anaphase promoting complex subunit 1
	Schisto_mansoni.Chr_W	31380464	31380941	TSS	Smp_153480	solute carrier family 12
	Schisto_mansoni.SC_0223	237419	237884	TSS	Smp_159890	carboxypeptidase d
	Schisto_mansoni.Chr_3.unplaced.SC_0077	29872	35737	TSS	Smp_160040	Abnormal long morphology protein 1
	Schisto_mansoni.Chr_1	23093504	23093979	TSS	Smp_160970	hypothetical protein
	Schisto_mansoni.Chr_1	51279957	51280469	TSS	Smp_161370	deoxynucleotidyltransferase terminal interacting
	Schisto_mansoni.Chr_1	25844210	25845022	TSS	Smp_161860	run domain Beclin 1 interacting and cysteine rich
	Schisto_mansoni.SC_0138	90963	92283	TSS	Smp_164450.1	tether containing UBX domain for GLUT4
	Schisto_mansoni.Chr_3	6256205	6256507	TSS	Smp_165770	nucleotide binding protein 1
	Schisto_mansoni.Chr_3.unplaced.SC_0174	515846	517102	TSS	Smp_165900	n2n2 dimethylguanosine tRNA methyltransferase
	Schisto_mansoni.SC_0357	29301	40000	TSS	Smp_170820	ATP binding cassette subfamily B (MDR:TAP)
	Schisto_mansoni.Chr_2.unplaced.SC_0193	56023	56552	TSS	Smp_171530	alpha parvin
	Schisto_mansoni.Chr_2.unplaced.SC_0193	110458	111496	TSS	Smp_171540.1	Ribosome binding protein 1
	Schisto_mansoni.Chr_1	28371135	28371702	TSS	Smp_172420	E3 ubiquitin protein ligase TRIP12
	Schisto_mansoni.Chr_4.unplaced.SC_0032	1535582	1539567	TSS	Smp_172590	family S10 unassigned peptidase (S10 family)
	Schisto_mansoni.Chr_W	14390470	14391611	TSS	Smp_175090.1	long-chain-fatty-acid--CoA ligase
	Schisto_mansoni.SC_0151	547304	550107	TSS	Smp_175120.1	Alpha (1,6) fucosyltransferase (EC 2
	Schisto_mansoni.Chr_2	32642153	32643254	TSS	Smp_180990	heat shock proteinB1 associated protein 1
					Smp_181000	ubiquitin thioesterase otu1
	Schisto_mansoni.Chr_4	30796304	30796786	TSS	Smp_198010	arf GAP with GTPase, ANK repeat and PH
	Schisto_mansoni.SC_0393	29272	30708	TSS	Smp_199670	cation transporting ATPase
	Schisto_mansoni.SC_0221	154482	156140	TSS	Smp_205070	hypothetical protein
	Schisto_mansoni.Chr_5	9174658	9175267	Exon	Smp_210130	C Jun amino terminal kinase interacting protein
	Schisto_mansoni.Chr_4.unplaced.SC_0032	1596005	1600000	TSS	Smp_211180	alpha(13)fucosyltransferase
H3K9Ac	Schisto_mansoni.Chr_3	1370338	1371116	Intergenic		
	Schisto_mansoni.Chr_4	4874988	4875379	Intergenic		
	Schisto_mansoni.Chr_6	1733270	1733941	Intergenic		
	Schisto_mansoni.Chr_1	57281606	57282474	Exon	smp_196370	potassium voltage gated channel subfamily H
	Schisto_mansoni.Chr_W	22785725	22787129	Intergenic		
H3K9me3	Schisto_mansoni.Chr_2	12214549	12215081	exon	Smp_014570.1	Saposin,IPR008139 Saposin
	Schisto_mansoni.Chr_1.unplaced.SC_0071	1531043	1531341	intron	Smp_075940	prohibitin protein WPH
	Schisto_mansoni.Chr_3	18283100	18284075	intron	Smp_084890	ectonucleotide pyrophosphatase:phosphodiesterase
	Schisto_mansoni.Chr_1	30851010	30852524	5' end	Smp_093980	hypothetical protein
	Schisto_mansoni.SC_0080	999585	1000073	3' end	Smp_133490	mitogen activated protein kinase 15
	Schisto_mansoni.Chr_7	6998392	6998741	intron	Smp_134560	serine:threonine protein kinase doa
	Schisto_mansoni.Chr_1	13972042	13972891	5' gene	Smp_141250	protein jagged 1
	Schisto_mansoni.Chr_2	11730991	11731254	exon	Smp_146570	hypothetical protein
	Schisto_mansoni.SC_0142	506493	506899	intron	Smp_158580	transmembrane protein 39a
	Schisto_mansoni.Chr_4	16199908	16200561	intron	Smp_160670	paired box protein pax 6
	Schisto_mansoni.Chr_1	24876731	24877219	intron	Smp_161680	phosphorylase B kinase beta, kpb
	Schisto_mansoni.Chr_W	252207	252845	exon	Smp_175930	ELAVprotein 2:3:4
H3K27me3	Schisto_mansoni.Chr_3	18098105	18099549	Intergenic		
	Schisto_mansoni.Chr_4	16750925	16752331	exon	Smp_171740	multidrug resistance associated protein 1
	Schisto_mansoni.Chr_W	9836427	9837293	Intergenic		
	Schisto_mansoni.Chr_4	23526845	23530256	Intergenic		
	Schisto_mansoni.Chr_1.unplaced.SC_0010	3860276	3860997	Intergenic		

Supplementary Table 2: List of repetitive elements where histone marks profiles were different between hycanthone resistant and control worms. Scaffolds and positions refers to the genome assembly v5.2 of *Schistosoma mansoni*.

Element ID	Comparison	Histone mark	Fold change (log2)	P-value	Est. copy number	localisation
R=1886	HycR < Control	H3K4me3	4.22	3.2e-02	22	dispersed
smr_947			0.48	9.6e-04	352	dispersed
Sm_SR1-E2		H3K9ac	0.42	2.7e-02	82	dispersed
R=9			0.38	7.6e-02	1001	dispersed
Sm_Saci3			0.36	9.3e-02	96	dispersed
R=869	HycR > Control		-0.55	8.4e-03	39	Chr.2:3960k-4050k, SC_0013: 2695k-2724k
Smr_7429			-0.5	2.2e-02	14	dispersed
Smr_2067			-0.5	7.8e-02	22	Chr.5:3010k-3060k
R=323		H3K27me3	-0.5	5.6e-02	32	Chr.3:22040k-22075k, SC_0703: whole, covering Smp_183020, SC_1309: whole, covering smp_186130
Smr_2801			-0.5	8.0e-02	10	Chr.5: 1261k-1277k
R=564			-0.48	2.7e-02	40	SC_0163:0-3k, entire contig SC_0397
R=592			-0.48	2.7e-02	14	Intermingled on SC_0147:20k-170k, Intermingled on SC_0163:526k-554k

CHAPITRE IV

EFFET DE LA SOUCHE D'HÔTE INTERMÉDIAIRE SUR L'ÉPIGÉNOME DE S. MANSONI

IV.1. MISE EN CONTEXTE

Dans cette partie, je me suis situé une fois de plus dans la continuité des travaux de Julie Lepesant. J'étais intéressé à continuer l'étude des variations épigénétiques chez un isolat de *S. mansoni* se développant dans des mollusques de souches différentes, mais sous une forme me permettant de discriminer les rôles respectifs de la génétique et de l'épigénétique. Je cherchais également à observer dans quelle mesure des modifications de la structure chromatinienne pouvaient être transmises d'un stade développemental à un autre, et éventuellement à une génération suivante.

Dans les précédentes expériences effectuées au laboratoire (**Figure 20**, page 52), une population de miracidia a été utilisée pour infecter deux isolats géographiquement distincts de mollusques (Lepesant 2012). Les mollusques étant infectés par plusieurs miracidia (issus de plusieurs parents), les cercaires collectées sont alors de plusieurs génotypes. Tel que mentionné précédemment, il y a peu de variabilité génétique dans la souche *SmBre* (aucune variabilité sur 14 marqueurs microsatellites neutres), mais elle n'est cependant pas inexistante. Du séquençage massif sur 20 individus de cette souche a permis de mesurer l'indice de diversité de Tajima à $1,8 \times 10^{-4}$ (Clément *et al.* 2013) (**Annexe 1**). Cette diversité est faible, mais il reste que de nombreux polymorphismes ont tout de même été identifiés. On ne peut donc pas exclure l'éventualité qu'un ou plusieurs polymorphismes soient impliqués dans les phénotypes observés au niveau des traits d'histoire de vie : prévalence, intensité, sex-ratio (Lepesant *et al.* 2013; Theron *et al.* 2014). Il existe aussi la possibilité que certaines structures chromatinienne soient associées à des génotypes précis (par des mécanismes en *cis*). Il est alors risqué d'affirmer que les différences épigénétiques observées entre les deux conditions sont l'effet de l'environnement et non le résultat de la sélection d'allèles ou epiallèles conférant un avantage dans l'infestation d'une des deux souches de mollusques. Nous avons aussi remarqué un biais fort et reproductible du sex-ratio dans la condition allopatrique. Un enrichissement (augmentation) du nombre de femelles apparaît systématiquement dans les échantillonnages de cercaires issues des mollusques *BgGua* (Guadeloupe, allopatrique) (Lepesant *et al.* 2013). Deux hypothèses peuvent l'expliquer: **(I)** les miracidia femelles ont une meilleure infectivité que les mâles vis à vis des mollusques allopatriques, ou **(II)** les sporocystes femelles produisent plus de cercaires dans les conditions allopatriques. Dans

les deux cas, à cause de la variation du sex-ratio, on peut imaginer que les différences de structure chromatinienne observées soient sexe-spécifiques et deviennent détectables en raison de la proportion plus élevées de cercaires femelles dans une condition.

J'ai décidé de modifier cette expérience, de manière à réduire les problèmes liés à la sélection d'allèles/épiallèles entre les différentes souches de mollusques et à la possibilité d'enrichissement d'épiallèles sexe-spécifiques dans une condition. Pour outrepasser ces biais, j'ai effectué des infections mono-miracidiale (un seul miracidium par mollusque) de *SmBre* dans un mollusque sympatrique (*BgBre*), suivi de greffes de sporocystes dans des mollusques allopatriques (*BgGua*) et d'autres mollusques sympatriques (*BgBre*). Cette technique, développée par mon laboratoire d'accueil (Jourdane & Theron 1980; Jourdane 1984) permet de prélever des sporocystes secondaires à partir d'un mollusque infecté, et de les transférer dans des mollusques sains. Les sporocystes secondaires (sp2) greffés régressent au stade de sporocystes primaires (sp1), se multiplient asexuellement et se transforment à nouveau en sporocystes secondaires (sp2), émettant des milliers de cercaires. Le gros avantage de cette technique est qu'elle permet d'obtenir de nombreux mollusques allopatriques et sympatriques infectés par un même génotype de *S. mansoni*. Cela offre l'opportunité d'obtenir des réplicats biologiques d'une même condition, avec un fond génétique identique. Les réplicats clonaux d'une même condition sont cruciaux car ils permettent de déterminer si les modifications de structures chromatiniennes sont induites aléatoirement, de manière ciblée, ou les deux. En travaillant sur les cercaires émises, et les vers adultes en étant issus, il devient possible d'étudier une héritabilité somatique de changements induits par l'environnement.

Ce projet était aussi l'occasion d'explorer de nouvelles marques d'histones chez *S. mansoni*. Deux nouveaux anticorps que j'ai testés, ciblant l'acétylation de la lysine 27 de l'histone 3 (H3K27ac) et la mono-méthylation de la lysine 20 de l'histone 4 (H4K20me1), ont donné satisfaction pour toutes les étapes du pipeline de validation (Cosseau & Grunau 2011). L'analyse des données de l'expérience sur l'hycanthone (chapitre III) et de l'expérience de Julie Lepesant sur l'hôte intermédiaire (**Figure 20**, page 52, Lepesant 2012) ont mis en évidence que la grande majorité des différences de structures chromatiniennes entre les conditions étaient observées sur les marques H3K4me3 et H3K27me3, qui ont été reconduites pour cette expérience.

IV.2. PUBLICATION ASSOCIÉE

Les résultats et conclusions de cette expérience ont été soumis sous la forme de l'article « *Frequency and mitotic heritability of epimutations in Schistosoma mansoni* » en septembre 2015 pour l'édition spéciale « *Epigenetic Studies in Ecology and Evolution* » de la revue **Molecular Ecology**. La version présentée dans ce manuscrit est une version révisée et acceptée pour publication le 24 janvier 2016 (post-soutenance). Les *Supplementary Figures* 1 à 3 et les *Supplementary Files* 3, 4, 6 et 8 de l'article sont présentés à sa suite. Les *Supplementary Files* 1, 2, 5 et 7 sont disponibles sur demande à david.roquis@yahoo.com.

Frequency and mitotic heritability of epimutations in *Schistosoma mansoni*

Journal:	<i>Molecular Ecology</i>
Manuscript ID:	MEC-15-1138.R2
Manuscript Type:	Special Issue
Date Submitted by the Author:	n/a
Complete List of Authors:	<p>Roquis, David; University of Perpignan, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE)</p> <p>Rognon, Anne; University of Perpignan, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE)</p> <p>Chaparro, Cristian; University of Perpignan, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE)</p> <p>Boissier, Jerome; University of Perpignan, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE); UMR5244, Laboratoire de Biologie et d'Ecologie Tropicale et Méditerranéenne</p> <p>Arancibia, Nathalie; University of Perpignan, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE)</p> <p>Cosseau, Céline; Univ.Perpignan Via Domitia, Ecologie et Evolution des Interactions, UMR 5244; University of Perpignan, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE); CNRS, Ecologie et Evolution des Interactions, UMR 5244</p> <p>Parrinello, Hugues; CNRS, MGX</p> <p>Grunau, Christoph; Univ.Perpignan Via Domitia, Ecologie et Evolution des Interactions, UMR 5244; CNRS, Ecologie et Evolution des Interactions, UMR 5244</p>
Keywords:	Coevolution, Contemporary Evolution, Evolutionary Theory, Molecular Evolution, Parasitology, Epigenetics

Frequency and mitotic heritability of epimutations in *Schistosoma mansoni*

David Roquis^{1,2}, Anne Rognon^{1,2}, Cristian Chapparo, Jerome Boissier^{1,2}, Nathalie Arancibia^{1,2}, Celine Cosseau^{1,2}, Hugues Parrinello³ & Christoph Grunau^{1,2}

¹ Université de Perpignan Via Domitia, Perpignan, F-66860, France,

² CNRS, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE), Perpignan, F-66860, France,

³ CNRS, GenomiX IBISA, Montpellier, France,

Abstract

Schistosoma mansoni is a parasitic platyhelminth responsible for intestinal bilharzia. It has a complex life cycle, infecting a freshwater snail of the *Biomphalaria* genus, and then a mammalian host. *S. mansoni* adapts rapidly to new (allopatric) strains of its intermediate host. To study the importance of epimutations in this process we infected sympatric and allopatric mollusk strains with parasite clones. ChIP-Seq was done on four histone modifications (H3K4me3, H3K27me3, H3K27ac and H4K20me1) in parallel with genome-wide DNA resequencing (i) on parasite larvae shed by the infected snails, and (ii) on adult worms that had developed from the larvae. No change in single nucleotide polymorphisms (SNP) and no mobilization of transposable elements (TE) was observed, but 58 – 105 copy number variations (CNV) within the parasite clones in different mollusks were detected. We also observed that the allopatric environment induces three types of chromatin structure changes: (i) host-induced changes on larvae epigenomes in 51 regions of the genome that are independent of the parasites' genetic background, (ii) spontaneous changes (not related to experimental condition or genotype of the parasite) at 64 locations, and (iii) 64 chromatin structure differences dependent on the parasite genotype. Up to 45 % of the spontaneous, but none of the host-induced chromatin structure changes were transmitted to adults. In our model, the environment induces epigenetic changes at specific loci but only spontaneous epimutations are mitotically heritable and have therefore the potential to contribute to transgenerational inheritance. We also show that CNV are the only source of genetic variation and occur at the same order of magnitude as epimutations.

Introduction

Adaptive evolution relies on the generation of heritable phenotypic variations. There is now an intense scientific discussion about the source of these variants (Laland *et al.* 2014). During the last couple of years, it has become clear that there are considerable non-genetic sources of heritable phenotypes (Jablonka *et al.* 1992; Jablonka & Lamb 2008; Danchin *et al.* 2011). In particular epigenetic modifications, that we define here tentatively as heritable changes in chromatin structure that can have an impact on gene expression, are in the center of a lively scientific debate (Kovalchuk 2011). Particularities of epimutations are (I) their metastability, *i.e.* changes in chromatin structure can be reversible and (II) their relations with the environment. In the genetic inheritance system, the environment acts solely as selective agent, but in the epigenetic system it could play a dual role, acting as an inducer of epimutations (Guerrero-Bosagna & Skinner 2011; Skinner *et al.* 2015) as well as applying selection on the resulting phenotypes. The question whether epimutations can generate new phenotypes is solved (Peaston & Whitelaw 2006; Nilsson & Skinner 2015). There are now many examples for phenotypes that are entirely or partially encoded by epialleles presenting changes in DNA methylation or histone modifications (Duncan *et al.* 2014). The role of the environment, however, is far from being comprehended. There is empirical support that epimutations are generated spontaneously, independently of environment but at a higher frequency than genetic mutations (van der Graaf *et al.* 2015). It was also argued that environmental stress could provoke adaptive and heritable changes through targeted épimutations. Others concluded that stress leads to an increase of the epimutation rate, thus representing an ephemeral, but inducible version of the genetic inheritance system, *i.e.* a way to induce reversible but possibly inheritable phenotypic changes (Jiang *et al.* 2014). The objective of the present study was to determine the frequency and nature of histone-based epimutations during a change of environment, and to measure their mitotic heritability. To address this question we used a parasite-host system. In these systems reciprocal selective forces, such as infectivity for the parasite and resistance for the host, are strong and adaptive evolution is expected to be fast (Thrall *et al.* 2007; Tellier *et al.* 2014).

The plathelminth *Schistosoma mansoni* is the causing agent of intestinal schistosomiasis, a human disease affecting more than 67 million people worldwide (Oliveira *et al.* 2004; King 2010). Its complex life cycle (**Supplementary Figure 1**) starts with eggs hatching in fresh water. A free-swimming larvae (miracidium) is released and will enter the tegument of the intermediate mollusk host of the *Biomphalaria* genus. Miracidia quickly transform into primary sporocysts, multiply asexually and differentiate into secondary sporocysts. They generate multiple cercariae, another free-swimming larvae, which leave the mollusk, infect the definitive mammalian host (rodent or primate), and transform into schistosomula. They follow a complex maturation and migration process in the mammalian body to reach the adult worm stage. Male and female worms form couples in the mesenteric veins, where eggs are released, go through the intestine wall and are excreted in the feces. *S. mansoni* can infect a wide range of intermediate and definitive host species (Martins 1957; DeJong *et al.* 2001). Changes in the intermediate host strain and species has a reproducible phenotypic impact on the parasite, and more specifically on its life history traits (sex-ratio, growth, prevalence and intensity of the infection) (Dias *et al.* 1988; Lepesant *et al.*

2013; Theron *et al.* 2014), but the underlying mechanisms are still unknown. Compatibility polymorphism is classically used to explain infectivity between *S. mansoni* and *B. glabrata* (Roger *et al.* 2008; Mitta *et al.* 2012). However, recent work has uncovered that epigenetic modifications could also play a role in the adaptation to new host strains (Knight *et al.* 2013). Knowing this, we were interested to find out what would be the impact of different host strains on the genome and epigenome of *S. mansoni*. In order to disentangle the impact of epimutations and mutations, we decided to work with genetically identical parasites and to let them develop in different snails. Miracidia, coming from sexual reproduction, are not clonal, and although it is relatively easy to infect snails with sibling miracidia with low genetic diversity between them, there will still be genetic diversity. To overcome this problem, we used sporocyst grafts (Jourdane & Theron 1980; Jourdane 1984). This technique allows, after a mono-miracidium infection, to extract secondary sporocysts and graft them into new mollusks. Secondary sporocysts revert to primary sporocysts once grafted, and then undergo the normal development until cercariae emission (Jourdane & Theron 1980; Jourdane 1984).

Here, we used a parasite strain with known genetic background (*SmBre* from Brazil; Clément *et al.* 2013) to infect a sympatric *Biomphalaria glabrata* strain (*BgBre* from Brazil) and an allopatric one (*BgGua* from Guadeloupe). Interactions between *SmBre* and the two mollusk strains have been previously characterized in great detail and the differences in life traits are known (Lepesant *et al.* 2013; Theron *et al.* 2014). Prevalence in the sympatric interaction (*SmBre* on *BgBre*) is 50-100%, and 30-85% in the allopatric interaction (*SmBre* on *BgGua*), depending on the number of miracidia used for infection (Theron *et al.* 2014). This shift in the environment does not represent a strong stress but has still measurable effects on the life history traits of the parasite (Lepesant *et al.* 2013). We sequenced the genomes of cercariae and adult parasites to search for somatic mutations, copy number variations and mobilization of transposable elements. DNA methylation is only present at a very low level in *S. mansoni* (Geyer *et al.* 2011), which allowed us to focused on chromatin marks based on histone modifications. We used chromatin immunoprecipitation followed by sequencing (ChIP-Seq). ChIP-Seq was done on cercariae emitted after grafts, but also on adult worms obtained by infecting mice with the same cercariae. The objective was to see if chromatin structure changes, potentially induced by development in different mollusks, are transmitted to the next developmental stage of the parasite. We selected four histone modifications (H3K4me3, H3K27ac, H3K27me3, H4K20me1), associated with different biological functions. H3K4me3 form narrow peaks (rarely larger than 2 kb wide) usually associated with the transcription start site (TSS) and promoters of transcriptionally competent genes (Berger 2007; Gu & Fire 2009; Kharchenko *et al.* 2010; Zentner & Henikoff 2013). H3K27me3 plays an important role in sequential expression and regulation of developmental genes, and is generally associated with repression of transcription (Kouzarides 2007; Trojer & Reinberg 2007; Bannister & Kouzarides 2011). Acetylation of H3K27 serves to discriminate active enhancers (where the mark is present) from poised ones (where the mark is absent) (Tie *et al.* 2009; Creighton *et al.* 2010; Liu *et al.* 2011; Shlyueva *et al.* 2014). It is described as a very punctate mark, similar to H3K4me3. The role of H4K20me1 is more controversial. Some authors report it being associated with distinct regions of silent chromatin (Sims *et al.* 2006; Vielle *et al.* 2012) in human and *C. elegans*, while others describe it as an essential mark for transcription activation (Wang *et al.* 2008; Li *et al.* 2011). It is

also involved in DNA repair and cell cycle (Jorgensen *et al.* 2013), and genome-wide increase of this mark impairs genome integrity (Schotta *et al.* 2008).

Our data show that changes in SNP and mobilization of TE do not contribute to genetic variation, but that CNV are relatively frequent. Roughly 10% of spontaneous CNV are mitotically heritable. In addition, we observed the occurrence of environment-directed (*i.e.* environment-dependent) and genotype-independent epimutations, as well as spontaneous epimutations (not related to experimental conditions or genotypes). However, while mitotic heritability of spontaneous epimutations can reach 45%, none of the environmentally targeted chromatin structure changes are transmitted to the adult worms. Our findings are therefore in favor of spontaneous epimutations as a possible source for heritable phenotypic variants that could be selected by the environment.

2. Materials and Method

The experimental procedure is schematically summarized in **Figure 1**.

Ethics statement

Housing, feeding and animal care followed the national ethical standards established in the writ of February 1st, 2013 (NOR: AGRG1238753A) setting the conditions for approval, planning and operation of establishments, breeders and suppliers of animals used for scientific purposes and controls. The French Ministère de l'Agriculture et de la Pêche and French Ministère de l'Éducation Nationale de la Recherche et de la Technologie provided permit A 66040 to our laboratory for experiments on animals and certificate for animal experimentation (authorization 007083, decree 87–848) for the experimenters.

Origin of the parasites and their hosts

The *Schistosoma mansoni* strain *SmBre* was originally sampled in Recife, Brazil in the 1960's. It was provided to our laboratory in 1975 by Pr. Y. Golvan (Faculté de Médecine de Paris – Saint Antoine) and has been since then maintained in its sympatric intermediate host strain *BgBre* of the mollusk *Biomphalaria glabrata* and *Mus musculus* (SWISS OF1) as definitive vertebrate host. *BgBre* (albino strain) is also from Recife and was also acquired in 1975. Albinism is of genetic origin and does not have an impact on the experimental design. The allopatric mollusk strain *BgGua* originates from the town of Dans Fond and arrived at our laboratory in 2005 (pers. comm. A. Théron). Mollusks were maintained as described in Boissier & Moné (2000).

Infection and sex determination

Thirty-two *B. glabrata* *BgBre* individuals were each infected with a single *SmBre* miracidium. Miracidia were freshly hatched from the liver of a single infected mouse. Miracidia hatching, as well as snails and mice infections, were performed as described in Boissier & Moné (2000). Mollusks were screened 30 days later for presence of parasites and cercariae shed by each positive *B. glabrata* were genotyped with sex markers (Portela *et al.* 2010). We chose to work only with female parasites, as it was previously shown that there is an increase in female cercariae production when *SmBre* develops in *BgGua* hosts (Lepesant *et al.* 2013), and also to avoid possible

bias caused by gender-specific epialleles. We selected two infected snails with a large number of sporocysts to be donors for sporocyst grafts. We named the female genotypes of the parasites in these two snails *Fa* and *Fb*.

Sporocyst graft

The grafts were done as detailed in Jourdan & Theron (1980) and Jourdan (1984). Large (10-12 mm) *BgBre* and *BgGua* mollusks were selected as receivers. They were anesthetized by incubating 4 hours in spring water mixed with sodium pentobarbital at a final concentration of 1.2 mg/mL. Donor mollusks had their shells carefully removed with tweezers and their digestive gland (where secondary sporocysts are located) was recovered. Using tweezers, explants of 1 mm³ containing 1-3 sporocysts were prepared and grafted within an hour. Explants were grafted in the cephalopedal sinus of receiver snails with a custom-made glass micro-needle attached to a 1 mL syringe. A small incision of the tegument above the genital pore was done to reach the cephalopedal sinus. Grafted snails were then kept in normal growing conditions. *Fa* sporocysts were grafted in 15 *BgGua* and 14 *BgBre*, and *Fb* sporocysts were transplanted in 17 *BgGua* and 15 *BgBre*.

Parasite preparation

Fifty days after transplantations, mollusks were screened under a stereomicroscope to look for secondary sporocysts. Uninfected mollusks were discarded. Cercariae were collected from infected snails 53 days after infection by gently pipetting from spring water, avoiding mucus and feces, and sedimented on ice. Water was removed and cercariae from each snail were stored separately at -80°C before being used for ChIP-Seq. As we were interested to see if epimutations induced by the development in different snail strains would be maintained in adults, we also infected mice with cercariae from some of the snails. For both *Fa* and *Fb* parasite genotype, we selected *BgBre* and *BgGua* snails emitting the highest amount of cercariae (> 5,000) to use for infection. For parasite genotype *Fa*, we had one *BgBre* and one *BgGua*, and for *Fb* we had two snails of each strain. For each snail, three mice were infected with 300 cercariae each, as described in Boissier & Moné (2000). Mice were sacrificed 42 days later by injection of sodium pentobarbital and female adult worms were recovered by retrograde perfusions of the hepatic portal system with citrate (7.5%) saline (8.5%) solution administrated through the left ventricle (Duvall & DeWitt 1967). Worms trapped in the liver or mesenteric system were collected after excising these organs. Worms from each mouse were collected separately and stored at -80°C.

Genomic DNA sequencing

Cercariae emitted from a mollusk infected with a single miracidium are generally considered genetically identical, but evidences exist that some genetic diversity can occur (Bayne & Grevelding 2003). We could not exclude that sporocyst grafts and development in an allopatric host were source of stress, which could activate mobilization of TE (Wijayawardena *et al.* 2015). Knowing this, we performed DNA pool sequencing (Pool-Seq) on adult worms developed from cercariae emitted after the grafts. We chose to sequence DNA of adults rather than cercariae because current tools for the analysis of Pool-Seq data do not perform well on large numbers of potentially different haploid genomes. It is not feasible to obtain sufficient amount of DNA for sequencing from less than 1,000 cercariae while 10 adults are enough. We selected four pools of

12, 9, 11 and 17 Fb genotype female adult worms. The first two pools originated from two different sympatric (*BgBre*) snails, while the last two are from a single allopatric (*BgGua*) snail, but isolated from two different mice. DNA was extracted following Beltran *et al.* (2008) and sequencing was done as paired-end 125 bp reads on an Illumina HiSeq 2500 at the Montpellier GenomiX facility (<http://mgx.cnrs.fr>). Sequencing was designed to have a genomic coverage of at least 3X per individual per sample.

Chromatin immunoprecipitation and sequencing (ChIP-Seq)

Native ChIP was performed following the protocol developed for *Schistosoma mansoni* by Cosseau *et al.* (2009) (also available online at http://methdb.univ-perp.fr/cgrunau/methods/native_chip_sm.html) on cercariae and adult worms. We used at least 5,000 cercariae or 20 adult worms. Immunoprecipitation was performed using the following antibodies: H3K4me3 (Millipore, cat# 04-745 lot# NG1680351, 4 µL per reaction), H3K27ac (Abcam, cat# ab4729 lot# GR-150-367-2, 4 µL per reaction), H3K7me3 (Diagenode, cat# C15410069 lot# A1821D/2, 8 µL per reaction), and H4K20me1 (Abcam, cat# ab9051 lot# GR158874-1, 4 µL per reaction). For each sample, we used a control without antibody to assess unspecific background (bound fraction) and input (unbound fraction). Inputs were used for normalization in all subsequent bioinformatics analyses. Antibodies were carefully tested for specificity as described in Cosseau & Grunau (2011) and saturating quantities were used. We had earlier shown that the abovementioned catalogue/lot numbers for anti-H3K4me3 and anti-H3K27me3 can be used for ChIP-Seq (Roquis *et al.* 2015). For H3K27ac and H4K20me1, we performed all the validation steps described by Cosseau & Grunau (2011), including a chromatin titration to assess specificity of the antibodies. Titrations, using increasing amount of antibodies over the same quantity of *S. mansoni* chromatin, proved that antibodies were indeed specific and able to immunoprecipitate all target chromatin using 4 µl of the antibody for 5,000 cercariae or 20 adult female worms (**Supplementary Figure 2**).

ChIP products were sequenced as single-end 50 bp reads on an Illumina HiSeq 2500 at the McGill University and Génome Québec Innovation Centre (<http://gginnovationcenter.com/index.aspx>). Briefly, fragmented DNA was quantified using a 2100 Bioanalyzer (Agilent Technologies). Libraries were generated robotically with 10 ng of fragmented DNA (range 100-300 bp) using the Kapa HTP Library Preparation Kit (Kapa Biosystems) as per the manufacturer's recommendations, except that adapters and PCR primers were diluted 100-fold. The size selection step was done after the PCR step and the number of PCR cycles was increased by six. Adapters and PCR primers were purchased from Integrated DNA Technologies whereas size selection has been performed on a Pippin Prep instrument (SAGE Biosciences Inc). Libraries were quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies) and the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (D-Mark). Average size fragment was determined using a LabChip GX (PerkinElmer) instrument. Cluster formation on the flow cell was performed using the cBot instrument (Illumina) with four indexed libraries per lane. Sequencing, in the form of 50-cycle single end reads, was performed on a HiSeq 2000/2500 (Illumina) running HCS software version 2.2.38. Demultiplexed FASTQ files were generated by allowing up to one mismatch in the index.

Quality control, alignment and peak calling

All data treatment was carried out under a local galaxy instance (Goecks *et al.* 2010)

(<http://bioinfo.univ-perp.fr>). Read quality was verified using the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). All samples had a sanger quality score above 30 for over 95% of their length.

Sequences were aligned to the *S. mansoni* reference genome (Protasio *et al.* 2012) with Bowtie2.1 (Langmead & Salzberg 2012) using parameters --end-to-end, --sensitive, --gbar 4 and filtered for unique match with samtools (Li *et al.* 2009) (samtools view -Sh -q *quality value 40-42* -F 0x0004 - | grep -v XS:i). SAM alignment files were converted into the bed format with pyicos (Althammer *et al.* 2011) and sorted with sortBed -i of the bedtools suite (Quinlan & Hall 2010). PCR duplicates were removed using the remduplicates function from Pyicos (Althammer *et al.* 2011), allowing a maximum of 4 duplicates per position. Random sampling using a custom Perl script was performed on each library to generate an equal number of reads per library for peak calling. Although random sampling is not mandatory, we found in previous work that it reduces false positives when searching for histone modification differences between samples. We took 7 million reads for H3K4me3, 23 millions for H3K27ac and H3K27me3, and 19 millions for H4K20me1. One of the samples, Cercariae Fb Gua 8, had a higher amount of PCR duplicates than the other samples, and we could only select 14 million reads for H3K27ac and 10 millions for H4K20me3. Ranger of Peakranger v1.16 (Feng *et al.* 2011) with P-value cutoff 0.0001, FDR cutoff 0.01, Read extension length 200, Smoothing bandwidth 99 and Delta 1 was used for peak identification. Input samples were used as negative control (-c). Wiggle files were visualized under IGV (Thorvaldsdottir *et al.* 2013). Reads mapping to repetitive regions were aligned on the 3,145 repetitive consensus sequences obtained from the same strain of *S. mansoni* (*SmBre*) in a previous study (Lepesant *et al.* 2012) (available at <http://methdb.univ-perp.fr/downloads/>) using Bowtie2.1 (Langmead & Salzberg 2012) evoking parameters --end-to-end, --sensitive, --gbar 4. In order to obtain identical effective library sizes for each sample, the same number of aligned reads was randomly sampled for each of the ChIP experiments (H3K4me3 = 2.5 million, H3K27ac = 7 million, H3K27me3 = 7 million, H4K20me1 = 8 million). After the counts had been normalized, they were compared between groups (allopatric vs sympatric, with three replicates; genotype Fa vs Fb, with two and four replicates) and between individual samples using the DESeq package (Anders & Huber 2010) of the R software. This package uses the negative binomial distribution to calculate the P-value for each element and for each comparison between two samples and to find elements that present significant differences in their enrichment levels.

Chromatin structure profiles

The analysis of the different histone modification combinations on the *S. mansoni* genome was performed using EpiCSeq (Mammanna & Chung 2015), using count normalization (*normalizecounts*) and binsize of 500bp (--binsize 500). We tested 4 to 16 different possible chromatin states and concluded that five states (--nstates 5) was the optimal number to avoid redundancy between states. Profiles were done on cercariae and adults, each time using the three datasets from the sympatric (BRE) condition as replicates for the bin count. In a similar way, we also compared sympatric (BRE) and allopatric (GUA) chromatin profiles for adults and cercariae. Annotation files used to generate chromatin profiles on genes, transcription start sites (TSS) and transcription end site (TES) were derived from an earlier established transcriptome (Roquis *et al.* 2015), by keeping a subset of genes longer than 5 kb (2,428 genes).

Comparative epigenomic analysis

For comparison of chromatin profiles of each mark, DiffBind v1.8.3 (Stark & Brown 2013) was used with FDR of 0.1. Comparisons were done for each histone modification between allopatric and sympatric conditions (3 replicates for each histone modification). Comparisons were also done for genotype Fa (two replicates) vs Fb (four replicates), and for each individual sample (no replicate). One library (cercariae Fb5 H3K4me3) was systematically excluded from these DiffBind analyses because it had more than 20% of PCR duplicates, indicating probably a too few starting material, thus reducing the number of replicates for this histone modification by one. Heatmaps were generated evoking `dba.plot Heatmap` (sample, contrast=1, correlations=FALSE) and PCA was done with `dba.plotPCA`(sample, contrast=1, th=.05). DiffBind (allopatric vs sympatric condition and genotype Fa vs Fb with minOverlap = 2) will test a peak if it is identified in at least two of the six total samples. It is the variance amongst all the replicates in each condition that drive the p-value calculated for the comparison. For sample-by-sample comparison, non-parametric bootstrapping or default value minOverlap = 1 was used for each sample. Both methods provided the same results. Differentially modified chromatin regions were saved as BED files for further analysis. Gene ontology enrichment analysis was done with BLAST2GO (Conesa *et al.* 2005). For analyses of significant differences in histone isoform enrichment in repetitive sequences, normalized aligned read counts were compared using the DESeq package of the R software (Anders & Huber 2010).

Comparative genomic analysis

For the detection of SNP frequency distortion, *i.e.* significant differences in allele frequency between sympatric and allopatric conditions, we used paired-end data for genotype Fb and single-end data corresponding to the ChIP-Seq input libraries for both Fa and Fb. All raw data were of excellent quality and no quality filter was necessary. Reads were aligned with Bowtie 2 (Langmead & Salzberg 2012) and default parameters. Unique reads were identified using the Bowtie 2 'XS:' flag and sorted with SamTools (Li *et al.* 2009). Alignment rates were between 54% and 59%, which is consistent to what has been previously achieved on *S. mansoni* (Roquis *et al.* 2015) and can be explained by the fact that we only align on unique sequences of the genome (47% of the genome is composed of repetitive elements (Lepesant *et al.* 2012). To homogenize effective library size, 56 million aligned reads were randomly subsampled for the paired-end libraries and used for further analysis. For single-end data, 24 million reads were subsampled. Pileup files were generated by SamTools invoking `-B -C 0 -d 250 -q 0 -Q 13`. For the comparison of allele frequencies we used PoPoolation2 (Kofler *et al.* 2011). Fisher's Exact Test was chosen to estimate the significance of allele frequency differences in samples for which we had no replicates (Fa), and the Cochran-Mantel-Haenszel (CMH) test was used to detect consistent allele frequency changes with replicates (Fb). Pileup files were converted to sync file with parameter `--min-qual 20`. And Fisher and CMH tests were used invoking `--min-count 2 --min-coverage 10 --max-coverage 200 --suppress-noninformative` for single-end data and `--min-count 12` for paired-end data. Copy-number variation (CNV) are another source of genetic variability and we used CNV-seq (Xie & Tammi 2009) to estimate the probability of CNV differences between the sympatric and the allopatric conditions. We used default parameters (log2 = 0.6, pvalue = 0.001, minw = 4) with a genome size of 364,541,798 bp. Results were converted into IGV or GWAS compatible formats and visualized under IGV. Differences in genomes Fa and Fb were detected using FreeBayes v0.9.20

(Garrison & Marth 2012) using parameters --theta 0.001 --ploidy 2 -J -m 1 -q 0 -R 0 -Y 0 -e 1000 -F 0.2 -C 2 --min-alternate-qsum 0 -G 1 --min-coverage 20. Resulting VCF files were compared with vcf-isec of the VCF tools (Danecek *et al.* 2011). CNV were searched for as described above.

The search for transposable element (TE) insertions was done by aligning the reads of the 4 Fb paired-end pools using bwa-mem from the BWA software (Li & Durbin 2009) to the reference genome of *S. mansoni*. Detection of the insertions and excision of elements was done using the TEMP software (Zhuang *et al.* 2014) with the required repeat and annotation files, parameters -x 10 and -c 12 for the insertion detection, and default parameters for the absence (excision) detection. The principle of this analysis is to use differences in mapping of forward and reverse reads to detect TE insertion events. BED files were generated from the output of TEMP and the files were sorted. In order to eliminate constitutive differences between the sequenced genomes and the reference genome we subtracted the positions that were common to all four genomes and visually inspected the remaining regions.

Comparative Transcriptomics

Parasites were used to infect sympatric and allopatric snail hosts for three generations as described in Lepesant *et al.* (2013). For RNA extraction, cDNA generation and sequencing, and comparative analysis with the TopHat2 and CuffLinks we followed Roquis *et al.* (2015). Briefly, reads were aligned to the genome and exon-intron structures are established individually for each sample. Cuffmerge was used to generate a consensus transcription annotation file in GFF format. We then used htseq-count with the “union” option (Anders *et al.* 2015) to obtain read-counts for each of the 18,844 loci (“genes”) identified by Cufflinks/Cuffmerge. A single count (“pseudocount”) was added to each readcount to avoid division by 0 problems. DESeq (Anders & Huber 2010) with default parameters was used for differential gene expression analysis. A gene ontology enrichment analysis was done with BLAST2GO (Conesa *et al.* 2005) and prediction of protein interaction networks was done with STRING10 (Szklarczyk *et al.* 2015).

Results

There are no significant differences in SNP or TE mobilization between SmBre in sympatric and allopatric condition

SNP frequencies were evaluated using Fisher’s exact test (for Fa without biological replicates) and the CMH test for Fb (with replicates). Both tests are robust but the CMH-test showed the highest power to identify interesting loci (Kofler & Schlotterer 2014). Following others (Barsh *et al.* 2012, Panagiotou *et al.* 2012, Kofler & Schlotterer 2014, Broer *et al.* 2013) we decided to establish a significant threshold for p-values at $p = 1 \times 10^{-5}$. Despite this relatively relaxed threshold, we detect no significant allele frequency distortion. We noted nevertheless that there are large blocks of 100 kb range in the genome (roughly one per chromosome) that show enrichment of non-significant AF differences between the sympatric and allopatric condition in all replicates, whatever genotype was used. We calculated nucleotide diversity between Fa and Fb using private SNP (*i.e.* with allele frequency 1). We found 1,570 SNP differences (nucleotide diversity 4.3×10^{-6} based on the total

genome length of 364,541,798 bp) and 72 CNV between cercariae Fa and Fb emitted from the sympatric host *BgBre*.

The analysis of transposable element insertion produced several thousand candidate regions which showed differences in the number of supporting reads or the detection of one of the two insertion borders with respect to the reference sequence. The elimination of the redundant positions and those pointing to opposite insertion borders showed that the detected positions were due to the difference between the sequenced genome and the reference genome. Once these differences were eliminated the analysis showed that no neoinsertions of elements could be detected.

There are significant differences in CNV between sympatric and allopatric SmBre

CNV-seq detected 58 - 105 CNV covering 0.1-0.2% of the genome in all comparisons (**Table 1a & 1b, Supplementary File 1**). Interestingly, some CNV covered or flanked the above-mentioned blocks of non-significant AF differences. To answer the question whether mitotically heritable CNV are systematically found in the allopatric condition, we compared these variants between Fa and Fb. CNV between sympatric Fa and Fb were excluded since they could represent simply differences to the reference genome. None of the regions *sensu stricto* fulfilled the condition to be a CNV in cercariae allopatric vs sympatric in both Fa and Fb, and CNV in adults allopatric vs sympatric in both Fa and Fb. Nevertheless, there is in one region (Schisto_mansoni.Chr_3:1,725,500-1,762,000) with a clustering of environment dependent CNV. Three single exon genes are predicted in the region (Smp_192030, Smp_101720 and Smp_201970) for which we did, however, not find transcripts in our RNA-Seq data. CNV between biological replicates in sympatric or allopatric condition do not show significant differences (**Table 1a & 1b**). CNV locations are provided in the supplementary material (**Supplementary File 1**).

Taken together, our data provide evidence that, upon exposure to the allopatric condition, point mutations and TE mobilization are below the detection limit. CNV occurs at appreciable frequency, but are not linked to a specific environmental condition (allopatric or sympatric). The frequency of CNV is not different between the sympatric and allopatric hosts. Twelve to fifteen percent of random CNV are maintained through development to adults. One region on chromosome 3 could be a CNV hot-spot activated upon exposure to the allopatric environment.

In two cases (Schisto_mansoni.Chr_2:717900-817000 and Schisto_mansoni.Chr_1:58049000-58059000), we found epimutations close to CNV, but not to SNP or transposable elements. After having established the degree and type of genetic mutations we attempted to characterize the epimutations.

The epigenome of S. mansoni comes in five main chromatin colors

We then performed ChIP-Seq experiments with the abovementioned antibodies for two genotypes (Fa and Fb) of cercariae that had developed in sympatric or allopatric snail hosts, and of the resulting female adults (a total of 60 libraries, details in **Supplementary File 2**). The analysis of the different histone modification combinations on the *S. mansoni* genome shows five main states

(also called “colors”; Filion *et al.* 2010), in both cercariae (**Figure 2**) and adults (**Supplementary Figure 3**). Both profiles are qualitatively similar. The most frequent histone mark combination (state 3 on **Figure 2** and **Supplementary Figure 3**) is a combination of H3K27me3-H3K27ac-H4K20me1, which covers 60% of the genome in cercariae and only 46% in adults. This state has a low presence around the transcription start site of genes (TSS) but is strongly enriched at the transcription end site (TES). The second most frequent state (state 2) is characterized by an enrichment of H3K27ac and H3K20me1, covering 17% of the genome of cercariae and 24% of adults. It is enriched upstream of the TSS, decreases drastically in the first kb past the TSS before being enriched again in the gene body and decreases once more near the 3’end of the gene. The absence of any of the four histone marks (state 5) covers 14% of the genome in both developmental stages of the parasite. It is not very common in gene bodies, but is particularly depleted around TSS and TES. State 4, characterized by a stronger presence of H3K27me3, combined with a lower amount of H3K27ac and H4K20me1 is very lowly represented around genes. Outside of genic regions, it can be found covering 5% of cercariae and 14% of adult genomes. The last combination of marks (state 1) presents a very strong enrichment in H3K4me3 located right after the TSS (3% of cercariae and 2.3% of adult genomes). We also compared chromatin state between sympatric and allopatric cercariae, as well as adults, but we did not observe any large-scale chromatin structure profile differences (data not shown). It should be noted that except for H3K27ac and H3K27me3, that are mutually exclusive, our approach can neither confirm nor exclude that histone modifications occur at the same histone or same nucleosome. All we can say is that there is co-occurrence at the same genomic site.

There are three types of chromatin changes upon exposure to allopatric hosts

For comparison of chromatin changes in the sympatric and allopatric conditions, one would ideally use these five chromatin colors. Unfortunately, no algorithm exists yet to do so. Instead, we performed what was feasible, *i.e.* a comparative analysis histone mark by histone mark for the four modifications using peak-calling (PeakRanger) and differential enrichment analysis (DiffBind). For the cercaria we found in average 8,378 peaks for H3K4me3, 15,866 peaks for H3K27ac, 20,784 peaks for H3K27me3, and 11,175 peaks for H4K20me1. In adults there were 11,967 peaks for H3K4me3, 9,1110 peaks for H3K27ac, 22,435 peaks for H3K27me3, and 17,403 peaks for H4K20me1 (**Supplementary File 3**).

Heatmaps generated by DiffBind provide a general overview of the chromatin landscape. They clearly show chromatin differences in cercariae clones exposed to the sympatric and allopatric host, and the resulting female adults (**Figure 3A** and **3B**). Principal component analysis, using genomic positions with differences on chromatin profiles, separated sympatric and allopatric samples extremely well, based on the number of differences in histone marks between samples (**Supplementary Figure 4A** and **4B**). We observed in cercariae three different types of chromatin changes depending on condition and genotype. In principle, to classify epimutations, one would fit a statistical model to test for a genotype effect, an environmental effect, and possibly the interaction, on peak presence and shape. In such cases one could detect a significant effect of (for instance) environment even if not all replicates showed the pattern consistently. To our

knowledge, such statistical model has not been developed and tested. We use here a simpler approach and classified epimutations into **(1)** host-induced epimutations, **(2)** spontaneous epimutations, and **(3)** genotype dependent epimutations. For convenience, we will also use the term “targeted epimutation” or “induced epimutations” for type 1 epimutation, and “random epimutations” for type 2. In other words, at a given developmental stage (cercariae or adults), type 1 epimutations must be present in all samples of the allopatric condition and absent in the sympatric (or vice versa). Type 2 epimutations are present in at least one sample of a condition or a genotype, but must not be present in all samples of this condition or genotype. Type 3 epimutations are present in all samples of one genotype (independently of the environmental condition) and absent in all samples of the other genotype. A potential caveat of this classification method lies in the relatively low number of replicates, leading potentially to an overestimation of type 1 epimutations. In addition, type 2 epimutations might not be stochastic in the strict sense. For instance, they might also be environment or genotype-specific but with incomplete “penetrance” among replicate samples.

For type 1 epimutations, using a FDR 0.1 (Diffbind’s default), we identified 131 differences for H4K20me2, 592 for H3K4me3, 346 for H3K27ac and 203 for H3K27me3, present in all replicates between sympatric and allopatric cercariae. All were verified by visual inspection, in which we looked for background changes that DiffBind had probably mistakenly identified as real differences. Basically, if a trained human eye (considering intensity, shape, and strength of neighboring peaks) cannot detect the differences spotted by Diffbind, we considered it as a false positive (see **Supplementary Figure 5** for examples). In general, validated epimutations corresponded to a p-value $\leq 1 \times 10^{-5}$ for H3K4me3 and $\leq 1 \times 10^{-15}$ for the other marks. It reduced the number of type 1 differences to a total of 51 genomic regions, among those all of them showed changes for H3K4me3, 11 for H3K27ac, 2 for H3K27me3, and 3 for H4K20me1. Examples are shown in **Figure 4A** and a detailed list of differences is available in **Supplementary File 4**. Interestingly, epigenetic diversity is higher in samples from the allopatric condition (**Figure 3A** and **Supplementary Figure 4A**). We found that 76,5% of induced epimutations are in transcribed regions previously identified with RNA-Seq in cercariae (Roquis *et al.* 2015).

As outlined above we defined spontaneous epimutations (type 2) as a chromatin structure change that occurs in one or more samples, but that is not consistently shared among replicates (**Figure 4B**). We detected 64 genomic locations where these random epimutations occurred, very often with multiple changes in the same region: 60 changes in H3K4me3, 9 in H3K27ac, 6 in H3K27me3, and 5 in H4K20me1. We identified that 79.7% of the differences were in transcribed regions (details available in **Supplementary File 4**). We detected these epimutations by comparing samples one by one on Diffbind results (**Supplementary File 5**), and verifying them through the same visual inspection as for type 1 epimutations. We noted that for the two replicates of genotype Fb in the allopatric condition, we systematically found several hundreds of random epimutation for the four histone isoforms. We believe that this is due to a technical issue leading to higher background in these ChIP-Seq libraries, which is consistent with our visual inspection.

We also looked for genotype related epimutations (type 3), *i.e.* chromatin structure changes that are identical in both environmental conditions, but different when comparing cercariae of genotype Fa and Fb (**Figure 4C**). After visual inspection (same as type 1 & 2), we observed a total of 64 genomic locations with chromatin structure differences between the two genotypes, often with two or more marks changing: 39 differences for H4K20me2, 41 for H3K4me3, 38 for H3K27ac and 42 for H3K27me3. 73.4% of the changes between Fa and Fb are in transcribed genomic regions. BED files are available as **Supplementary File 5**.

Spontaneous epimutations, but not host-induced epimutations in cercariae are inherited to adults

By infecting mice with some of the cercariae collected from each sympatric and allopatric snail, we produced adult worms and investigated if the different types of epimutations we had spotted in cercariae were transmitted to the next developmental stage. None of the type 1 chromatin structure changes, shared by all replicates, were maintained in adults. However, Diffbind heatmaps (with the exception of H3K4me3) still show differences between adults developed from cercariae that have experienced either the sympatric or allopatric condition (**Figure 3B**), regardless of the genotype (Fa or Fb). With the exception of H3K27me3, PCA of adults separate adults by the environmental condition they experienced as cercariae (**Supplementary Figure 4A**). While host-induced epimutations were not inherited *sensu stricto* from cercariae to adults, the developmental trajectory was altered, resulting in different chromatin profiles between adults from allopatric or sympatric origin (**Figure 3A and 3B; Supplementary Figure 4A and 4B**).

Since we had replicates for genetically identical clones Fb, we investigated whether spontaneous epimutation (type 2) would be transmitted to the corresponding adults. Indeed, 45% of the sites with epimutations have the same chromatin structure in adults: 29 differences in H3K4me3, 3 in H3K27ac, 5 in H3K27me3 but none in H4K20me1 were heritable *sensu stricto*, *i.e.* they had occurred in the sporocysts and were maintained in cercariae and adults.

We found genotype-associated epimutations to be most mitotically heritable. Roughly 77% of the genomic locations with epimutations between Fa and Fb cercariae were found to be identical in adults. It corresponds to 31 changes in H3K4me3, 23 in H3K27ac, 24 in H3K27me3 and in 30 H4K20me1.

We then used the average number of peaks detected by PeakRanger (**Supplementary File 3**) to calculate frequency and heritability of histonic epimutations. Environmentally triggered epimutations (type 1) occur at a frequency of 4.2×10^{-3} (H3K4me3), 1.2×10^{-3} (H3K27ac), 8.9×10^{-5} (H3K27me3), and 1.7×10^{-4} (H4K20me1). Random epimutations (type 2) are observed in the same order of magnitude: 5×10^{-3} (H3K4me3), 1×10^{-3} (H3K27ac), 2.7×10^{-4} (H3K27me3), and 2.9×10^{-3} (H4K20me1). As mentioned above, heritability of type 1 epimutations is 0 (reversibility = 1). In contrast, heritability of type 2 epimutations reaches up to 0.83. In summary, frequency and heritability of histonic epimutations depends on the chemical nature of the modification and their type. Mean frequency is in the 10^{-3} order of magnitude.

Based on their locations, epimutations for the different histone marks were manually compiled into a single table for further analysis (**Supplementary File 4**).

Negligible chromatin structure changes occur around repetitive elements

Mapping our ChIP-Seq library over the 3,145 identified repetitive sequences of *S. mansoni* genome and analyzing the read counts between samples, we did not find any differences present in all replicates of an environmental condition in cercariae or adults. Only when we compared environmental conditions within a single genotype, or between genotypes in the same conditions, we detected changes in 21 repetitive elements (17 of which being transposable elements). Only a single histone mark was modified in these 21 repetitive elements. None of the histone marks were systematically changed and we found no heritability of the changes from cercariae to adults (**Supplementary File 6**).

Chromatin structure changes are not associated with specific pathways or molecular/biological functions

After having established that chromatin architecture changes upon exposure to a different environment and that this occurs primarily in regions of the genome that contain genes, we wondered whether a phenotype was associated with these epimutations. GO enrichment analysis was done for all the transcribed regions of the genome where chromatin structure changes were detected (host-induced, spontaneous and genotype-associated epimutations). No specific pathway or enrichment in molecular or biological functions was observed in any type of comparison. STRING analysis showed that Smp_016630 (Trimeric G-protein alpha o subunit), Smp_141980 (cAMP-specific 3,5-cyclic phosphodiesterase) and Smp_147260 (Ras-like) could potentially interact in a signaling pathway. All three genes have environmentally triggered complex epimutations (**Supplementary File 4**).

Next, we exposed miracidia for three generations to the allopatric host, extracted RNA from the resulting cercariae and compared their RNA profile to cercariae produced in sympatric conditions. We had earlier shown that transcription is stalled in cercariae (Roquis *et al.* 2015) and we expected only small differences in RNA profiles. We compared the RNA levels of genes at or close to the type 1 epimutations using the three generations as biological replicates (**Supplementary File 7**). Differences in RNA levels of epimutated genes were moderate with a few exceptions: Smp_130050 codes probably for a PDZ domain containing RING finger protein and its RNA is 3-5 fold overrepresented in allopatric cercariae. Smp_151290 codes for a putative multidrug and toxin extrusion protein 2 and its RNA is 2.5 times enriched in cercariae from allopatric hosts. Smp_015630 (Glutamate-gated chloride channel subunit) is threefold underrepresented in allopatric cercariae.

We also identified genes with strong differences *i.e.* adjusted p-value < 0.01, and/or 10 fold overrepresentation in one of the conditions (**Supplementary File 8**). For these genes we verified chromatin structure in cercariae and adults. None of the genes were in the regions that we had qualified as epimutated. However, it is noteworthy that 80% (50 of 62) of these genes do not

contain a canonical H3K4me3 peak at the TSS. This is two times higher than the genome average. It could therefore be that with the antibody combination we used we did not capture the chromatin structure of these genes. We also noticed subtle differences in the extent and position of chromatin states 1 to 5 (**Supplementary Figure 6**). We would have considered this as random noise but in some cases the chromatin state profile was faithfully inherited to the adults, suggesting that these small differences are stable and might have an impact on gene expression.

Discussion

In this experiment we set out to investigate the impact of the environment on the genome and epigenome of *S. mansoni* larvae. The environment could be (i) a selective agent, (ii) could induce specific mutations and/or epimutations, and (iii) could change mutation and/or epimutation rate, or (iv) might have no effect. The two *S. mansoni* genotypes we used in this study (Fa and Fb) come from the same inbred population. Nucleotide diversity is 0.0004% based on SNP, but they show relatively strong differences on the level of CNV (0.2% of the genome). We also found that they bear epigenetic differences. Here, we considered genotype-associated epimutations as chromatin structure changes present within a genotype in both allopatric and sympatric condition, but absent in the other genotype. We observed 64 of these differences ($\text{FDR} \leq 0.1$, $p \leq 1 \times 10^{-5}$ and after visual inspection). The majority (73.4%) of epimutations are in transcribed regions in cercariae, which is similar to what we observed with spontaneous and induced epimutations. In at least two striking cases that we investigated, genotype associated epimutations were linked to CNV. In the case of clone Fa, there is an important CNV difference between cercariae emitted from sympatric and allopatric snails, and could be a result of the environment, although without replicates for this genotype we cannot assess it without doubt. It has been previously observed (Skinner *et al.* 2015) that environmentally induced epimutations can lead to genetic mutation and CNV. We noticed that the CNV were clustered (**Supplementary File 1**). Centromere positions on *S. mansoni* genome are not mapped to the genome, and it is also possible that these CNV clusters are located in pericentromeric regions, which are hotspots of segmental duplications and prone to generate CNV (Makino *et al.* 2013). Apparition of these CNV could also possibly explain some of the genetic heterogeneity between *S. mansoni* clonal sporocysts observed before (Bayne & Grevelding 2003).

We then looked at environmentally triggered epigenetic changes, *i.e.* epimutations that occur in all the samples of a given mollusk condition. We found a total of 51 genomic regions, where at least one histone mark is different in all three cercariae samples emitted from sympatric and allopatric mollusks. Changes were often subtle (peaks of low intensities compared to neighbor peaks). As we used whole organisms for ChIP-Seq, there is the possibility that these changes only reflect a small subset of cells or tissues within cercariae. No gene pathway or gene ontology term enrichments were uncovered in genes spanning induced epimutations, and none of these epimutations were transmitted to adults. This type of ephemeral epigenetic change in response to an environmental stimulus is reminiscent of what was found for instance in smokers and former smokers where roughly a dozen CpG sites change their methylation status but revert progressively to normal when smoking was abandoned (Wan *et al.* 2012; Elliott *et al.* 2014).

No major chromatin remodeling was observed in the 3,145 known repetitive sequences of the *S. mansoni* genome, and although some of the epigenetic differences are located in transposable elements, we did not detect new insertion sites. Transposable elements movement is known to generate new genetic, epigenetic and phenotypic diversity in other works (Biemont 2010; Fedoroff 2012; Holoch & Moazed 2015), but they do not play a major role in this study.

Subsequently, we investigated the presence of spontaneous (type 2) epimutations through the genome of cercariae. We first detected type 2 epimutations through comparison of replicates of clone Fb. Sixty-four sites on the genome showed random epimutation, and 45% of them maintained their epimutated chromatin structure at the adult stage. Our data suggest that a change of the environment has an impact on epimutations since chromatin profiles of cercariae emitted from BgGua (and the resulting adults) present more differences between samples (as observed from the PCA on **Supplementary Figure 4A and 4B**) than chromatin profiles of cercariae and adults emitted from BgBre. We had shown before that *S. mansoni* worms exposed transiently to hycanthone, an antelmintic drug, possess 57 days later a characteristic chromatin profile different from unexposed worms (Roquis *et al.* 2013). At that time, we were not able to tell if the drug directly induced the different profiles, or if it acted as a selection agent, killing most of the parasites in the population, except those with a pre-existing distinct chromatin structure. In light of our new findings we are now in favor of the hypothesis that random epimutations had occurred prior to the drug exposure, and had produced in some individuals a resistant phenotype to the treatment.

The average frequency of histone modification based epimutations in our system is about 10^{-3} . This is one order of magnitude higher than what was reported for DNA methylation based epimutations in CpG pairs (Schmitz *et al.* 2011; van der Graaf *et al.* 2015). In this issue, Kronholm & Collins (2015) underline the importance of stability of epimutations for adaptive evolution. We show that heritability (*i.e.* stability) depends strongly on the type of epimutation and that for most of the detected epimutations reversion is actually above the threshold of 10^{-1} that Kronholm and Collins define as the reversion rate that limits the adaptive effect of epimutations. In other words, our data support the view that in *S. mansoni* epimutations do not contribute to adaptive walks, unless they provide a strong fitness gain.

Schistosoma mansoni proves to be an interesting model for this kind of epigenetic studies, due to its ecological/healthcare importance, the possibility to generate clones, and different phenotypically distinct developmental stages allowing to observe mitotic inheritance of epimutations. However, one of the caveats of this organism is its small size, making it impossible at the moment to perform ChIP-Seq on specific tissues or individuals (we needed at least 5,000 cercariae for our experiments). It means that we can only observe epimutations that are present in most of the cells and most individuals in the pools. Tissue-specific epimutations, as well as rare epimutations within the clone population, cannot be detected. Conversely, this also means that the chromatin structure changes we observed concern the majority of cells and individuals in our pools, and hence are more likely to have a significant adaptive impact. We also keep in mind that our sample size (three for each environmental condition and developmental stage) is small, and

more replicates, from the same and new genotypes, need to be done in order to strengthen our findings.

Conclusion

In conclusion, our findings indicate that in *S. mansoni* sporocysts, there are two major sources of mitotically heritable variation: CNV and spontaneous epimutations. This means that at least in our biological model (i) epimutations occur at appreciable frequency and since they are overrepresented in coding regions they can have an impact on gene expression, and (ii) that CNV occur at the same order of magnitude as epimutations, potentially contributing as much as the latter to phenotypic variation. We could not fully answer the question whether or not environmental changes increase epimutations. That means that at least for *S. mansoni* theoretical models in which epimutations serve to produce phenotypic variants that explore the fitness landscape for invariant genotypes (e.g. Klironomos *et al.* 2013) are fully applicable but future models must include CNV and must take into account the high reversion rate of histone-based epimutations.

Acknowledgement and data availability

Authors are grateful to Rory Stark, Aaron Taudt, Maria Colomé-Tatche and Frank Johannes for helpful discussions. Data for ChIP-Seq, Pool-Seq and RNA-Seq are available as fastq files BioProject Accession PRJNA236156 at the NCBI SRA. Chromatin profiles and transcripts can be visualized at our GBrowse instance (Stein *et al.* 2002) (<http://genome.univ-perp.fr>). Funding was provided by the French National Research Agency ANR, project ANR-10-BLAN-1720 (EPIGEVOL) and the GenEpi program of the UMR5244/IHPE (<http://ihpe.univ-perp.fr/en/ihpe-transversal-epigenetics/>).

Figure legends

Figure 1: Experimental procedure for the sporocyst grafts. Brazilian mollusks (*BgBre*) were infected with a single miracidium of the sympatric strain of *S. mansoni* (*SmBre*). Thirty days later, secondary sporocysts were surgically removed from the infected mollusks and grafted in several sympatric (*BgBre*) and guadeloupean allopatric (*BgGua*) mollusks. Emitted cercariae were collected for chromatin immunoprecipitations followed by sequencing (ChIP-Seq) and infection of mice to obtain adult worms. Adult worms were recovered for ChIP-Seq and paired-end DNA sequencing. The experiment was performed twice, with two different genotypes of the parasite (Fa and Fb). In Fb genotype, two replicates were done in sympatric and allopatric conditions.

Figure 2: Five different chromatin states (“colors”) as identified by EpiCSeq in cercariae. (A) Composition in histone marks of the five chromatin states. Color code of the states is used in all other parts of this figure. (B) Proportion of the genome covered by each state in cercariae. Enrichment for each state around gene bodies (C), at transcription start site (TSS) (D) and transcription end site (TES) (E) of cercariae. Enrichment profiles were done on 2,428 genes longer than 5 kb. Results for adults are very similar and shown in **Supplementary Figure 2**.

Figure 3: Heatmaps for the four studied histone marks in cercariae (A) and adults (B). Only sites with statistically significant differences (epimutations) were used. Red bars highlight conserved changes. Red snail symbols represent sympatric host *B. glabrata* BgBre, blue snail symbols stand for allopatric snail host BgGua. Cercariae sample GuaFb5 for H3K4me3 was excluded from analysis (see methods).

Figure 4: Typical examples of epimutations. (A) environmentally dependent epimutations (green frame) in H3K4me3 (blue profile) and H3K27ac (red profile): there are differences in all three samples Fa and two replicates Fb between sympatric and allopatric cercariae. (B) Random epimutations in H3K4me3, present in cercariae in three samples and transmitted to adult worms in corresponding three samples. (C) genotype-associated epimutations in H3K4me3, H3K27ac, H3K27me3 and H4K20me1 between genotypes Fa and Fb in sympatric condition. The track Ref. transcriptome *S. mansoni* displays transcripts (exons as blocks and introns as arrows) deduced from RNA-Seq data in cercariae.

Supplementary Figure 1: Life cycle of the parasite *Schistoma mansoni*. The endoparasites infect two hosts, (1) a mollusk intermediate host (*Biomphalaria* genus) in which it multiplies asexually as sporocysts and then cercariae, and (2) a vertebrate definitive host (primate or rodent) in which adult worms reproduce sexually. Two types of free-swimming larvae, miracidia and cercariae, are the infective form for the mollusk and vertebrate hosts, respectively. Eggs laid by adult worm couples in the mesenteric vein of the vertebrate host can go through the intestine wall to be excreted in the feces, and hatch in contact with freshwater.

Supplementary Figure 2: ChIP titration, using the same amount of *S. mansoni* chromatin (equivalent to 20 adult worms) and an increasing amount of antibody. Percentage of input recovery of chromatin was measured by qPCR on two housekeeping genes, alpha tubulin and 28S, as described in (Cosseau *et al.* 2009). Titrations were done to measure the efficiency of immunoprecipitation for antibody H3K27ac Abcam cat# ab4729 lot# GR-150-367-2 (A and B) and H4K20me1 Abcam, cat# ab9051 lot# GR158874-1 (B and C).

Supplementary Figure 3: Five different chromatin states ("colors") as identified by EpiCSeq in adults. (A) Composition in histone marks of the five chromatin states. Color code of the states is used in all other parts of this figure. (B) Proportion of the genome covered by each state in adults. Enrichment for each state around gene bodies (C), at transcription start site (TSS) (D) and transcription end site (TES) (E) of adult worms. Enrichment profiles were done on 2,428 genes longer than 5 kb.

Supplementary Figure 4: Principal Component Analysis (PCA) for the four studied histone marks in cercariae (A) and adults (B). Only sites with statistically significant differences (epimutations) were used. Blue circles (Bre) represent cercariae and adults from sympatric conditions (*SmBre* in *BgBre*), magenta (Gua) the allopatric conditions (*SmBre* in *BgGua*). Cercariae sample GuaFb5 for H3K4me3 was excluded from analysis (see methods).

Supplementary Figure 5: Examples of two regions (**A** and **B**) that were identified as epimutations (displayed by the red boxes) by DiffBind but did not pass the visual inspection.

Supplementary Figure 6: Example of a region that was not identified as epimutated by DiffBind but shows differences based on EpiCSeq chromatin “colors”, judged by visual inspection. The associated gene is differentially expressed and chromatin states are transmitted to adults.

References

- Althammer S, Gonzalez-Vallinas J, Ballare C, Beato M, Eyra E (2011) Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics*, **27**, 3333–3340.
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology*, **11**, R106.
- Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
- Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Nature Publishing Group*, **21**, 381–395.
- Barsh GS, Copenhaver GP, Gibson G, Williams SM (2012) Guidelines for genome-wide association studies. *PLoS Genetics*, **8**, e1002812.
- Bayne CJ, Greveling CG (2003) Cloning of *Schistosoma mansoni* sporocysts in vitro and detection of genetic heterogeneity among individuals within clones. *Journal of Parasitology*, **89**, 1056–1060.
- Beltran S, Galinier R, Allienne JF, Boissier J (2008) Cheap, rapid and efficient DNA extraction method to perform multilocus microsatellite genotyping on all *Schistosoma mansoni* stages. *Memorias do Instituto Oswaldo Cruz*, **103**, 501–503.
- Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature*, **447**, 407–412.
- Biemont C (2010) A brief history of the status of transposable elements: from Junk DNA to major players in evolution. *Genetics*, **186**, 1085–1093.
- Boissier J, Moné H (2000) Experimental observations on the sex ratio of adult *Schistosoma mansoni*, with comments on the natural male bias. *Parasitology*, **121** (Pt 4), 379–383.
- Broer L, Lill C, Schuur M *et al.* (2013) Distinguishing true from false positives in genomic studies: p values. *European Journal of Epidemiology*, **28**, 131–138.
- Clément JAJ, Toulza E, Gautier M *et al.* (2013) Private selective sweeps identified from next-generation pool-sequencing reveal convergent pathways under selection in two inbred *Schistosoma mansoni* strains (AR Jex, Ed.). *PLoS Neglected Tropical Diseases*, **7**, e2591.
- Conesa A, Gotz S, Garcia-Gomez JM *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.
- Cosseau C, Grunau C (2011) Native chromatin immunoprecipitation. *Methods in molecular biology (Clifton, N.J.)*, **791**, 195–212.

- 789 Cosseau C, Azzi A, Smith K *et al.* (2009) Native chromatin immunoprecipitation (N-ChIP) and ChIP-
790 Seq of *Schistosoma mansoni*: Critical experimental parameters. *Molecular and Biochemical*
791 *Parasitology*, **166**, 70–76.
- 792 Creighton MP, Cheng AW, Welstead GG *et al.* (2010) Histone H3K27ac separates active from
793 poised enhancers and predicts developmental state. *Proceedings of the National Academy of*
794 *Sciences*, **107**, 21931–21936.
- 795 Danchin E, Charmantier A, Champagne FA *et al.* (2011) Beyond DNA: integrating inclusive
796 inheritance into an extended theory of evolution. *Nature Reviews Genetics*, **12**, 475–486.
- 797 Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*,
798 **27**, 2156–2158.
- 799 DeJong RJ, Morgan JA, Paraense WL *et al.* (2001) Evolutionary relationships and biogeography of
800 *Biomphalaria* (*Gastropoda: Planorbidae*) with implications regarding its role as host of the
801 human bloodfluke, *Schistosoma mansoni*. *Molecular Biology and Evolution*, **18**, 2225–2239.
- 802 Dias LC de S, Bruce JI, Coles GC (1988) Strain variation in the infectivity of *Schistosoma mansoni* for
803 *Biomphalaria glabrata*. *Revista do Instituto de Medicina Tropical de São Paulo*, **30**, 86–90.
- 804 Duncan EJ, Gluckman PD, Dearden PK (2014) Epigenetics, plasticity, and evolution: How do we link
805 epigenetic change to phenotype? *Journal of Experimental Zoology Part B: Molecular and*
806 *Developmental Evolution*, **322**, 208–220.
- 807 Duvall RH, DeWitt WB (1967) An improved perfusion technique for recovering adult schistosomes
808 from laboratory animals. *The American journal of tropical medicine and hygiene*, **16**, 483–486.
- 809 Elliott HR, Tillin T, McArdle WL *et al.* (2014) Differences in smoking associated DNA methylation
810 patterns in South Asians and Europeans. **6**, 1–10.
- 811 Fedoroff NV (2012) Transposable elements, epigenetics, and genome evolution. *Science (New*
812 *York, N.Y.)*, **338**, 758–767.
- 813 Feng X, Grossman R, Stein L (2011) PeakRanger: A cloud-enabled peak caller for ChIP-seq data.
814 *BMC Bioinformatics*, **12**, 139.
- 815 Filion GJ, van Bommel JG, Braunschweig U *et al.* (2010) Systematic protein location mapping
816 reveals five principal chromatin types in *Drosophila* cells. *Cell*, **143**, 212–224.
- 817 Garrison E, Marth G (2012) *Haplotype-based variant detection from short-read sequencing*. 2012.
- 818 Geyer KK, pez CMRIGLO, Chalmers IW *et al.* (2011) Cytosine methylation regulates oviposition in
819 the pathogenic blood fluke *Schistosoma mansoni*. *Nature Communications*, **2**, 424–10.
- 820 Goecks J, Nekrutenko A, Taylor J, Galaxy Team T (2010) Galaxy: a comprehensive approach for
821 supporting accessible, reproducible, and transparent computational research in the life
822 sciences. *Genome Biology*, **11**, R86.
- 823 Gu SG, Fire A (2009) Partitioning the *C. elegans* genome by nucleosome modification, occupancy,
824 and positioning. *Chromosoma*, **119**, 73–87.
- 825 Guerrero-Bosagna C, Skinner MK (2011) Environmentally induced epigenetic transgenerational
826 inheritance of phenotype and disease. *MOLECULAR AND CELLULAR ENDOCRINOLOGY*, **354**, 1–
827 6.
- 828 Holoch D, Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. *Nature*
829 *Reviews Genetics*, **16**, 71–84.
- 830 Jablonka E, Lamb MJ (2008) Soft inheritance: challenging the modern synthesis. *Genetics and*

- 831 *Molecular Biology*, **31**, 389–395.
- 832 Jablonka E, Lachmann M, Lamb MJ (1992) Evidence, mechanisms and models for the inheritance
833 of acquired characters. *Journal of Theoretical Biology*, **158**, 245–268.
- 834 Jiang C, Mithani A, Belfield EJ *et al.* (2014) Environmentally responsive genome-wide accumulation
835 of de novo *Arabidopsis thaliana* mutations and epimutations. *Genome Research*, **24**, 1821–
836 1829.
- 837 Jorgensen S, Schotta G, Sorensen CS (2013) Histone H4 Lysine 20 methylation: key player in
838 epigenetic regulation of genomic integrity. *Nucleic Acids Research*, **41**, 2797–2806.
- 839 Jourdan J (1984) Maintenance of a male and a female clone of *Schistosoma mansoni* by
840 microsurgical transplantation of sporocysts. Reliability of the method. *Annales de*
841 *parasitologie humaine et comparée*, **59**, 361–367.
- 842 Jourdan J, Theron A (1980) *Schistosoma mansoni*: cloning by microsurgical transplantation of
843 sporocysts. *Experimental parasitology*, **50**, 349–357.
- 844 Kharchenko PV, Alekseyenko AA, Schwartz YB *et al.* (2010) Comprehensive analysis of the
845 chromatin landscape in *Drosophila melanogaster*. *Nature*, **471**, 480–485.
- 846 King CH (2010) Parasites and poverty: The case of schistosomiasis. *Acta Tropica*, **113**, 95–104.
- 847 Klironomos FD, Berg J, Collins S (2013) How epigenetic mutations can affect genetic evolution:
848 Model and mechanism. **35**, 571–578.
- 849 Knight M, Arican-Goktas HD, Ittiprasert W *et al.* (2013) Schistosomes and snails: a molecular
850 encounter. *Frontiers in genetics*, **5**, 230–230.
- 851 Kofler R, Schlotterer C (2014) A guide for the design of evolve and resequencing studies. *Molecular*
852 *Biology and Evolution*, **31**, 474–483.
- 853 Kofler R, Pandey RV, Schlotterer C (2011) PoPoolation2: identifying differentiation between
854 populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, **27**, 3435–
855 3436.
- 856 Kouzarides T (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.
- 857 Kovalchuk I (2011) Transgenerational epigenetic inheritance in animals. *Frontiers in genetics*, **3**,
858 76–76.
- 859 Kronholm I, Collins S (2015) Epigenetic mutations can both help and hinder adaptive evolution.
860 *Molecular Ecology*, n/a–n/a.
- 861 Laland K, ULLER T, Feldman M, Sterelny K, Müller GB (2014) Does evolutionary theory need a
862 rethink? *Nature*.
- 863 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**,
864 357–359.
- 865 Lepesant JMJ, Boissier J, Climent D, Cosseau C, Grunau C (2013) Female biased sex-ratio in
866 *Schistosoma mansoni* after exposure to an allopatric intermediate host strain of *Biomphalaria*
867 *glabrata*. *Experimental parasitology*, **135**, 350–356.
- 868 Lepesant JMJ, Roquis D, Emans R *et al.* (2012) Combination of de novo assembly of massive
869 sequencing reads with classical repeat prediction improves identification of repetitive
870 sequences in *Schistosoma mansoni*. *Experimental parasitology*, **130**, 470–474.
- 871 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform.

- 872 *Bioinformatics*, **25**, 1754–1760.
- 873 Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools.
874 *Bioinformatics*, **25**, 2078–2079.
- 875 Li Z, Nie F, Wang S, Li L (2011) Histone H4 Lys 20 monomethylation by histone methylase SET8
876 mediates Wnt target gene activation
- 877 Liu T, Rechtsteiner A, Egelhofer TA *et al.* (2011) Broad chromosomal domains of histone
878 modification patterns in *C. elegans*. *Genome Research*, **21**, 227–236.
- 879 Makino T, McLysaght A, Kawata M (2013) Genome-wide deserts for copy number variation in
880 vertebrates. *Nature Communications*, **4**, 1–10.
- 881 Mammana A, Chung H-R (2015) Chromatin segmentation based on a probabilistic model for read
882 counts explains a large portion of the epigenome. *Genome Biology*, 1–12.
- 883 Martins AV (1957) Non-human vertebrate hosts of *Schistosoma haematobium* and *Schistosoma*
884 *mansoni*. *Bulletin of The World Health Organization*, **18**, 931–944.
- 885 Mitta G, Adema CM, Gourbal B, Loker ES, Theron A (2012) Compatibility polymorphism in
886 snail/schistosome interactions: From field to theory to molecular mechanisms. *Developmental*
887 *& Comparative Immunology*, **37**, 1–8.
- 888 Nilsson EE, Skinner MK (2015) Environmentally induced epigenetic transgenerational inheritance
889 of disease susceptibility. *Translational Research*, **165**, 12–17.
- 890 Oliveira G, Rodrigues NB, Romanha AJ, Bahia D (2004) Genome and genomics of schistosomes.
891 *Canadian Journal of Zoology*, **82**, 375–390.
- 892 Panagiotou OA, Ioannidis JPA, for the Genome-Wide Significance Project (2012) What should the
893 genome-wide significance threshold be? Empirical replication of borderline genetic
894 associations. *International journal of epidemiology*, **41**, 273–286.
- 895 Peaston AE, Whitelaw E (2006) Epigenetics and phenotypic variation in mammals. *Mammalian*
896 *genome : official journal of the International Mammalian Genome Society*, **17**, 365–374.
- 897 Portela J, Grunau C, Cosseau CL *et al.* (2010) Whole-genome in-silico subtractive hybridization
898 (WISH) - using massive sequencing for the identification of unique and repetitive sex-specific
899 sequences: the example of *Schistosoma mansoni*. *BMC Genomics*, **11**, 387.
- 900 Protasio AV, Tsai IJ, Babbage A *et al.* (2012) A systematically improved high quality genome and
901 transcriptome of the human blood fluke *Schistosoma mansoni* (KF Hoffmann, Ed.). *PLoS*
902 *Neglected Tropical Diseases*, **6**, e1455.
- 903 Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features.
904 *Bioinformatics*, **26**, 841–842.
- 905 Roger E, Mitta G, Moné Y *et al.* (2008) Molecular determinants of compatibility polymorphism in
906 the *Biomphalaria glabrata*/*Schistosoma mansoni* model: New candidates identified by a global
907 comparative proteomics approach. *Molecular and Biochemical Parasitology*, **157**, 205–216.
- 908 Roquis D, Lepesant MJM, Picard MAL *et al.* (2015) The epigenome of *Schistosoma mansoni* provides
909 insight about how cercariae poise transcription until infection (SC Oliveira, Ed.). *PLoS*
910 *Neglected Tropical Diseases*, **9**, e0003853.
- 911 Roquis D, Lepesant MJM, Villafan E *et al.* (2013) Exposure to hycanthone alters chromatin structure
912 around specific gene functions and specific repeats in *Schistosoma mansoni*. *Frontiers in*
913 *genetics*, **5**, 207–207.

- 914 Schmitz RJ, Schultz MD, Lewsey MG *et al.* (2011) Transgenerational Epigenetic Instability Is a
915 Source of Novel Methylation Variants. *Science (New York, N.Y.)*, **334**, 369–373.
- 916 Schotta G, Sengupta R, Kubicek S *et al.* (2008) A chromatin-wide transition to H4K20
917 monomethylation impairs genome integrity and programmed DNA rearrangements in the
918 mouse. *Genes & Development*, **22**, 2048–2061.
- 919 Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to genome-
920 wide predictions. *Nature Reviews Genetics*, **15**, 272–286.
- 921 Sims JK, Houston SI, Magazinnik T, Rice JC (2006) A trans-tail histone code defined by
922 monomethylated H4 Lys-20 and H3 Lys-9 demarcates distinct regions of silent chromatin. *The*
923 *Journal of biological chemistry*, **281**, 12760–12766.
- 924 Skinner MK, Guerrero-Bosagna C, Haque MM (2015) Environmentally induced epigenetic
925 transgenerational inheritance of sperm epimutations promote genetic mutations. *Epigenetics*,
926 **10**, 762–771.
- 927 Stark R, Brown G (2013) DiffBind: differential binding analysis of ChIP-Seq peak data.
- 928 Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva,
929 A, Lewis S (2002) The Generic Genome Browser: a building block for a model organism system
930 database. *Genome Research*, **12**, 1599–1610.
- 931 Szklarczyk D, Franceschini A, Wyder S *et al.* (2015) STRING v10: protein-protein interaction
932 networks, integrated over the tree of life. *Nucleic Acids Research*, **43**, D447–D452.
- 933 Tellier A, Moreno-Gómez S, Stephan W (2014) Speed of adaptation and genomic footprints of
934 host-parasite coevolution under arms race and trench warfare dynamics. *Evolution*, **68**, 2211–
935 2224.
- 936 Theron A, Rognon A, Gourbal B, Mitta G (2014) Multi-parasite host susceptibility and multi-host
937 parasite infectivity: a new approach of the *Biomphalaria glabrata*/*Schistosoma mansoni*
938 compatibility polymorphism. *Infection, Genetics and Evolution*, **26**, 80–88.
- 939 Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-
940 performance genomics data visualization and exploration. *Briefings in Bioinformatics*, **14**, 178–
941 192.
- 942 Thrall PH, Hochberg ME, Burdon JJ, Bever JD (2007) Coevolution of symbiotic mutualists and
943 parasites in a community context. *Trends in Ecology & Evolution*, **22**, 120–126.
- 944 Tie F, Banerjee R, Stratton CA *et al.* (2009) CBP-mediated acetylation of histone H3 lysine 27
945 antagonizes *Drosophila* Polycomb silencing. *Development*, **136**, 3131–3141.
- 946 Trojer P, Reinberg D (2007) Facultative heterochromatin: is there a distinctive molecular
947 signature? *Molecular Cell*, **28**, 1–13.
- 948 van der Graaf A, Wardenaar R, Neumann DA *et al.* (2015) Rate, spectrum, and evolutionary
949 dynamics of spontaneous epimutations. *Proceedings of the National Academy of Sciences*,
950 **112**, 6676–6681.
- 951 Vielle A, Lang J, Dong Y *et al.* (2012) H4K20me1 contributes to downregulation of X-linked genes
952 for *C. elegans* dosage compensation (WG Kelly, Ed.). *PLoS Genetics*, **8**, e1002933.
- 953 Wan ES, Qiu W, Baccarelli A *et al.* (2012) Cigarette smoking behaviors and time since quitting are
954 associated with differential DNA methylation across the human genome. *Human Molecular*
955 *Genetics*, **21**, 3073–3082.

- 956 Wang Z, Zang C, Rosenfeld JA *et al.* (2008) Combinatorial patterns of histone acetylations and
957 methylations in the human genome. *Nature Genetics*, **40**, 897–903.
- 958 Wijayawardena BK, DeWoody JA, Minchella DJ (2015) The genomic proliferation of transposable
959 elements in colonizing populations: *Schistosoma mansoni* in the new world. *Genetica*, 1–12.
- 960 Xie C, Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-
961 throughput sequencing. *BMC Bioinformatics*, **10**, 80.
- 962 Zentner GE, Henikoff S (2013) Regulation of nucleosome dynamics by histone modifications.
963 *Nature Publishing Group*, **20**, 259–266.
- 964 Zhuang J, Wang J, Theurkauf W, Weng Z (2014) TEMP: a computational method for analyzing
965 transposable element polymorphism in populations. *Nucleic Acids Research*, **42**, 6826–6838.
- 966
- 967 **Data Accessibility**
- 968 RNA-Seq and ChIP-Seq rawdata are available as fastq files under study accession number
969 SRP035609 (BioProject Accession number PRJNA236156) at the SRA of the NCBI.

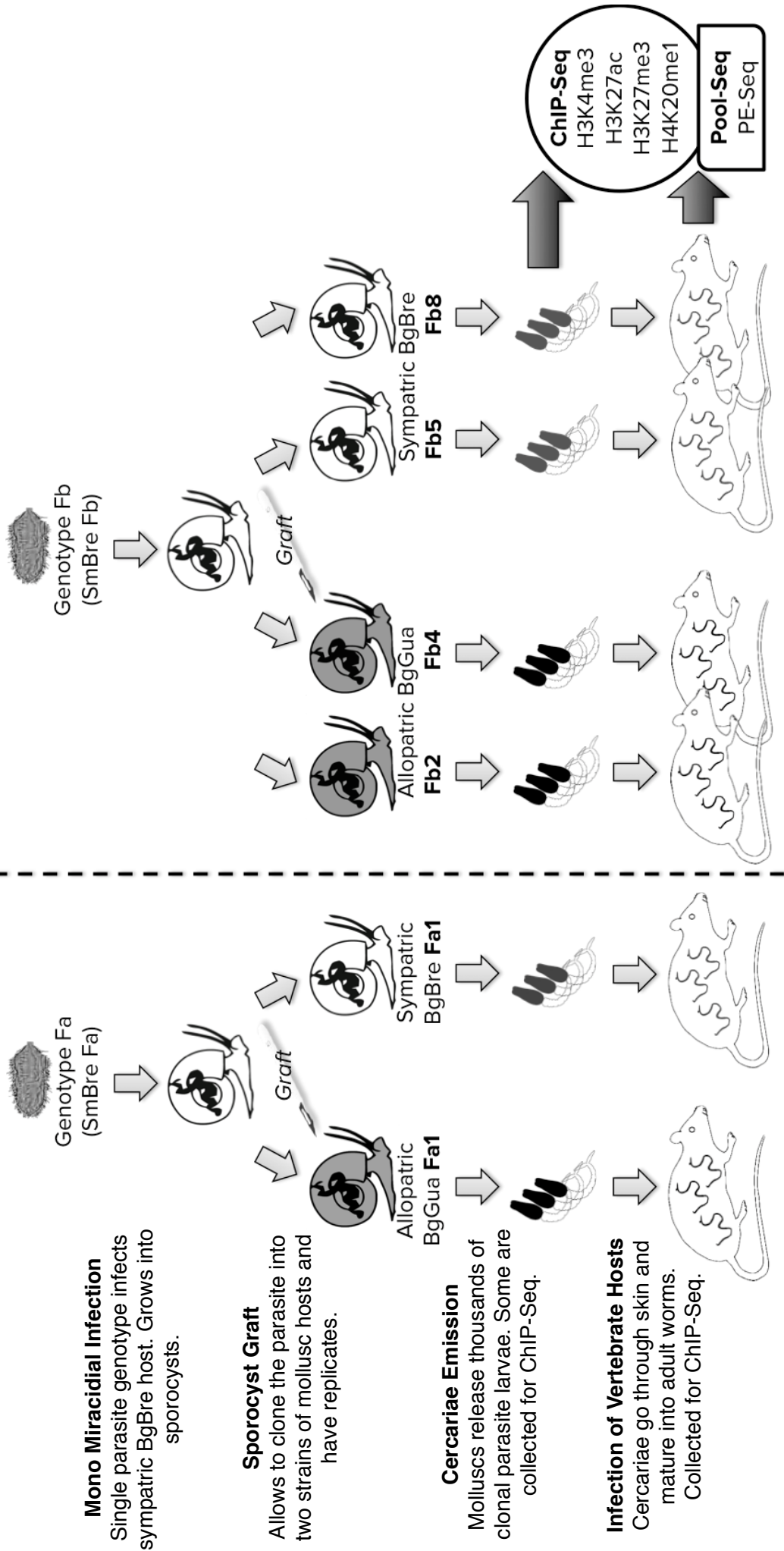
Table 1a: Amount of copy number variations (CNV) as a function of condition

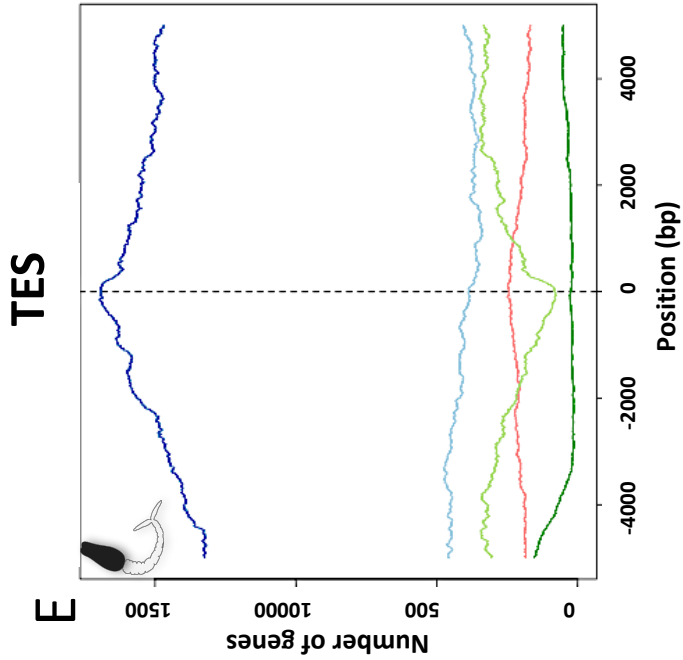
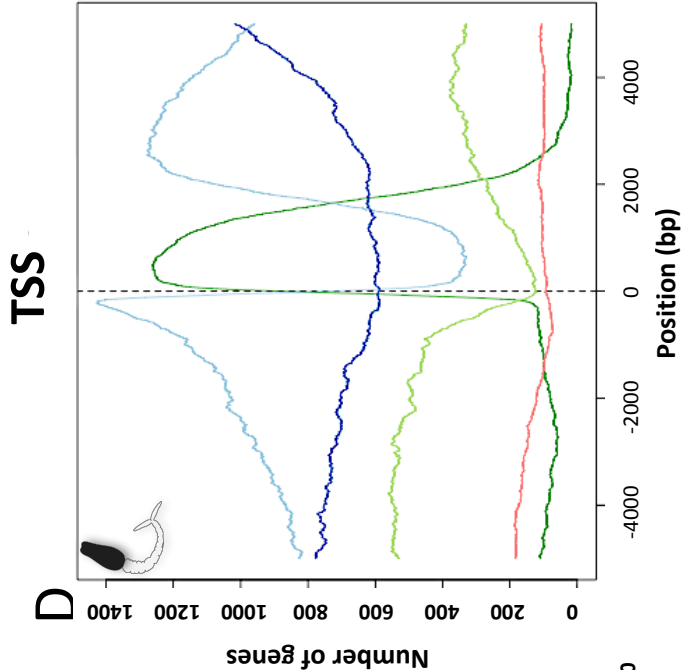
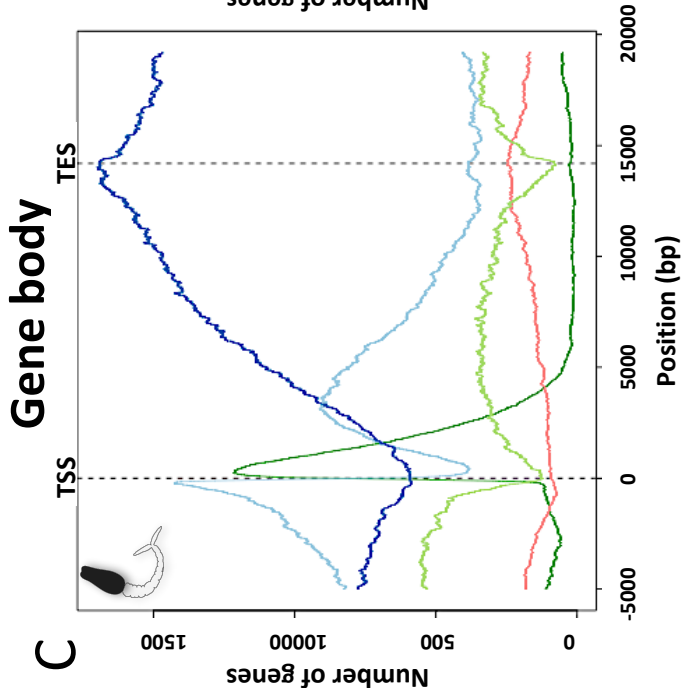
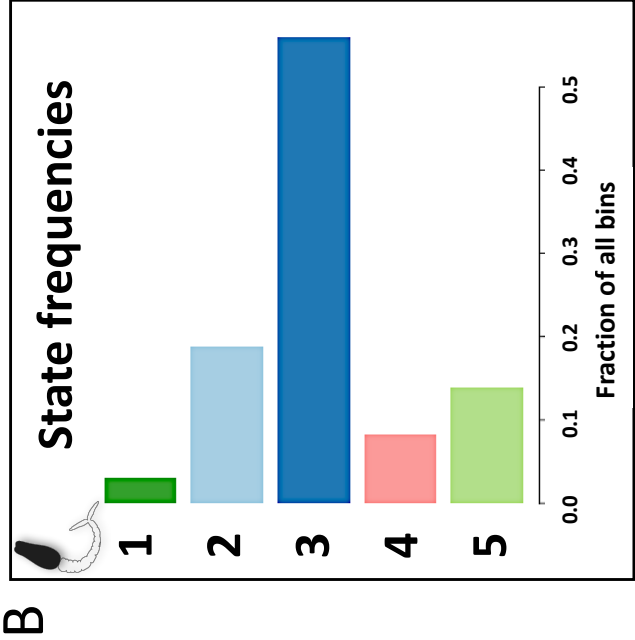
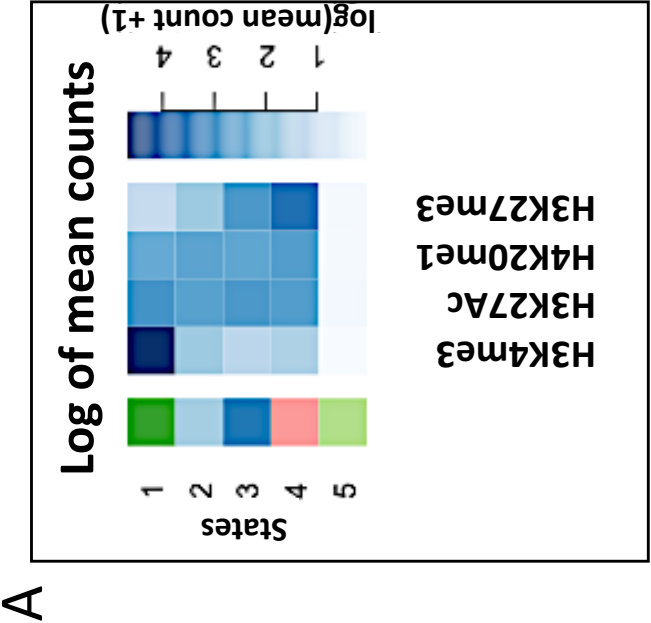
CNV number / heritable CNV (in %)	Cercariae <i>SmBre</i> Fa from <i>BgBre</i>	Adults <i>SmBre</i> Fa from <i>BgBre</i>	Cercariae <i>SmBre</i> Fb(2) from <i>BgBre</i>	Adults <i>SmBre</i> Fb(2) from <i>BgBre</i>
Cercariae <i>SmBre</i> Fa from <i>BgGua</i>	105			
Adults <i>SmBre</i> Fa from <i>BgGua</i>		97 / 12 (12.4%)		
Cercariae <i>SmBre</i> Fb(5) from <i>BgGua</i>			64	
Adults <i>SmBRE</i> Fb(5) from <i>BgGUA</i>				58 / 9 (15.5%)

Table 1b: Amount of copy number variations (CNV) as a function of genotype

Total CNV number	Cercariae <i>SmBre</i> Fb(2) from <i>BgBre</i>	Adults <i>SmBre</i> Fb(2) from <i>BgBre</i>	Cercariae <i>SmBre</i> Fb(5) from <i>BgGua</i>	Adults <i>SmBre</i> Fb(5) from <i>BgGua</i>
Cercariae <i>SmBre</i> Fb(4) from <i>BgBre</i>	67			
Adults <i>SmBre</i> Fb(4) from <i>BgBre</i>		42		
Cercariae <i>SmBre</i> Fb(8) from <i>BgGua</i>			72	
Adults <i>SmBre</i> Fb(8) from <i>BgGua</i>				63

Figure 1





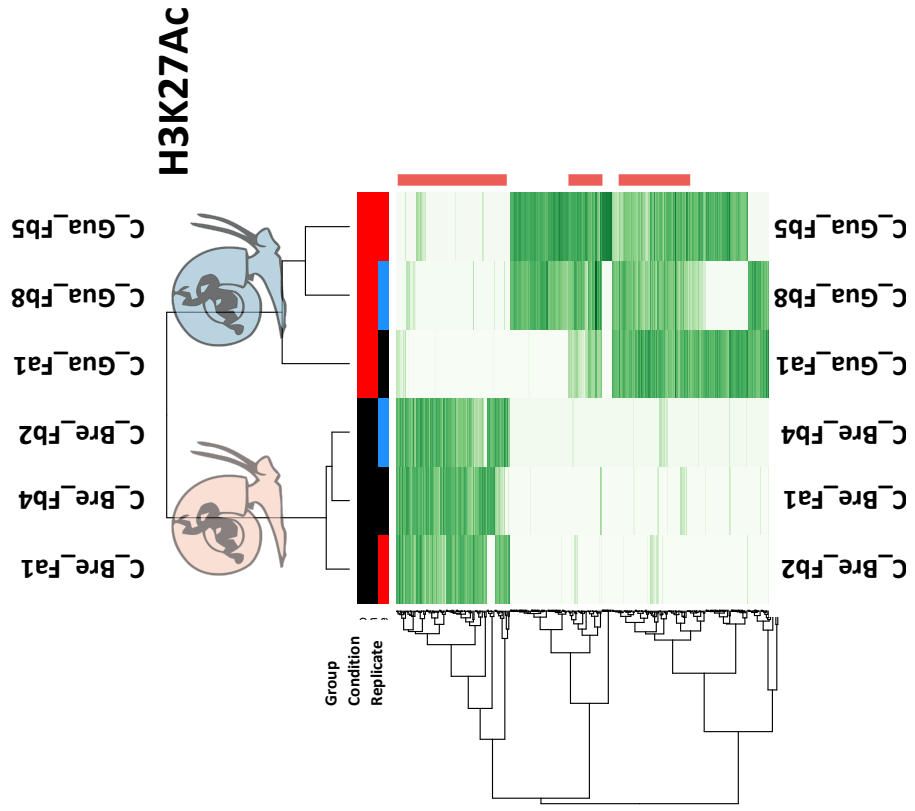
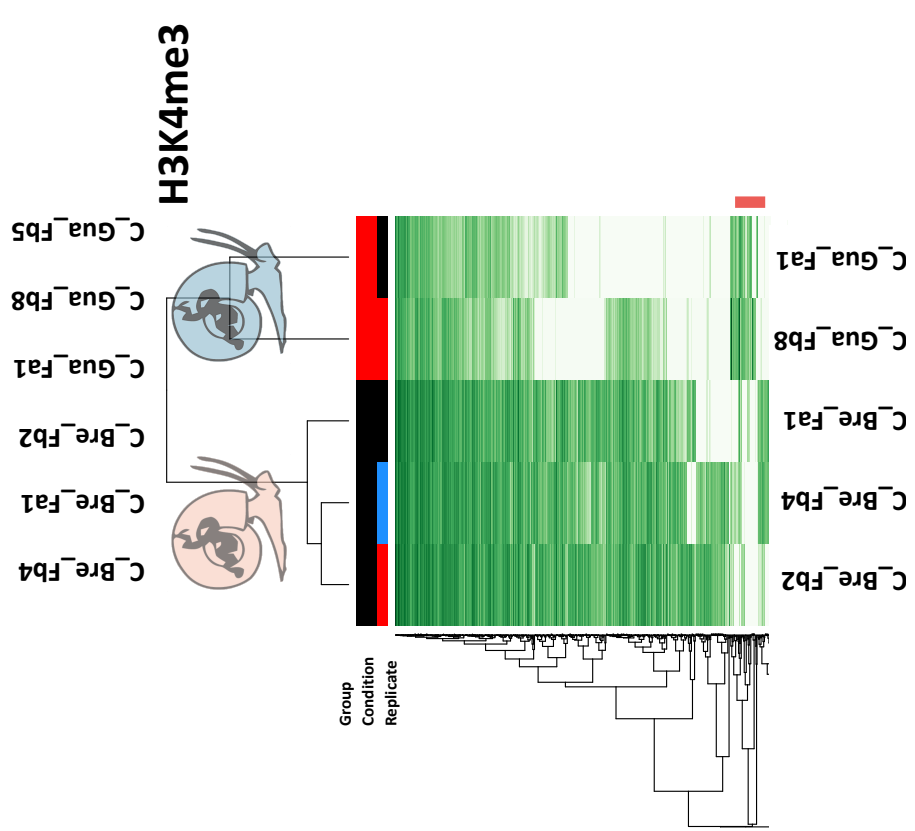
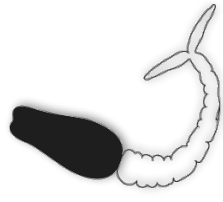
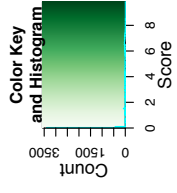
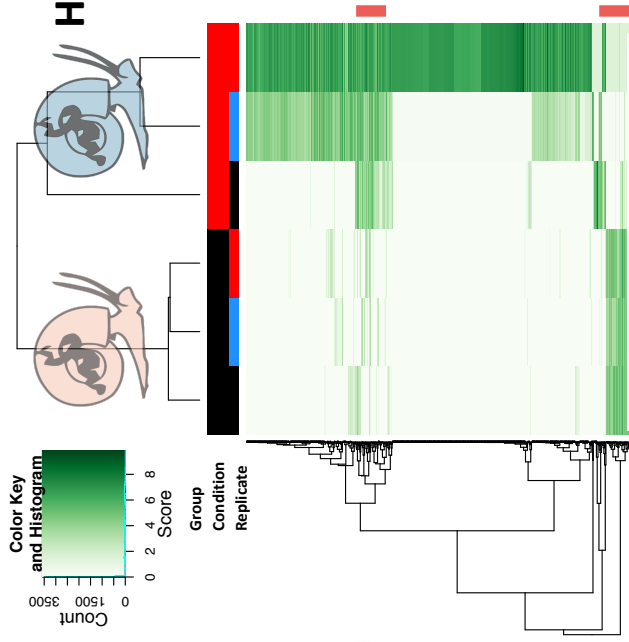
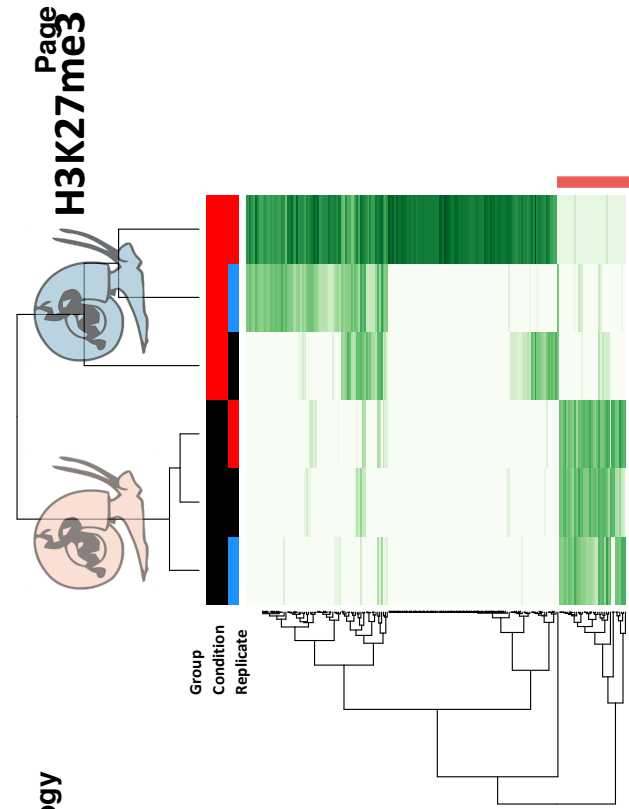
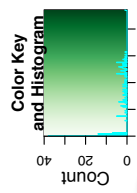
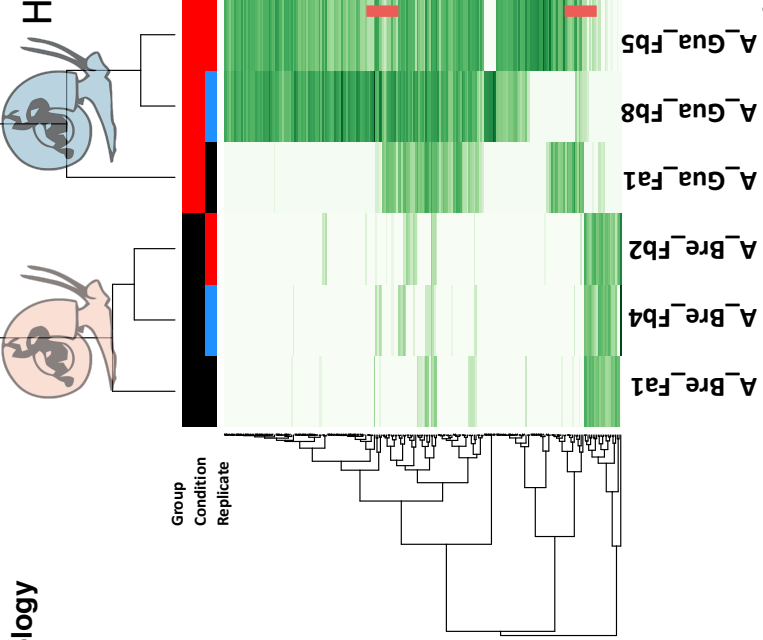


Figure 3

H4K20me1 Ecology

H3K27me3



H3K27AC

H3K4me3

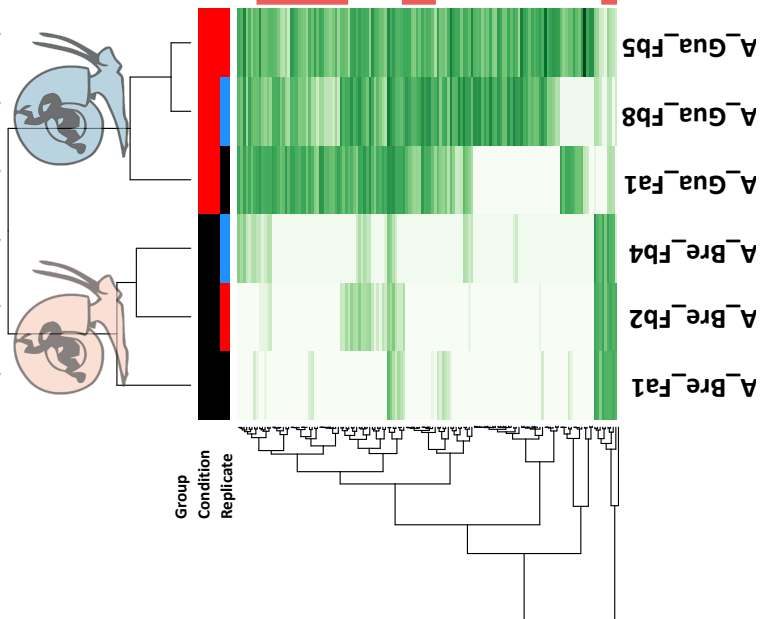
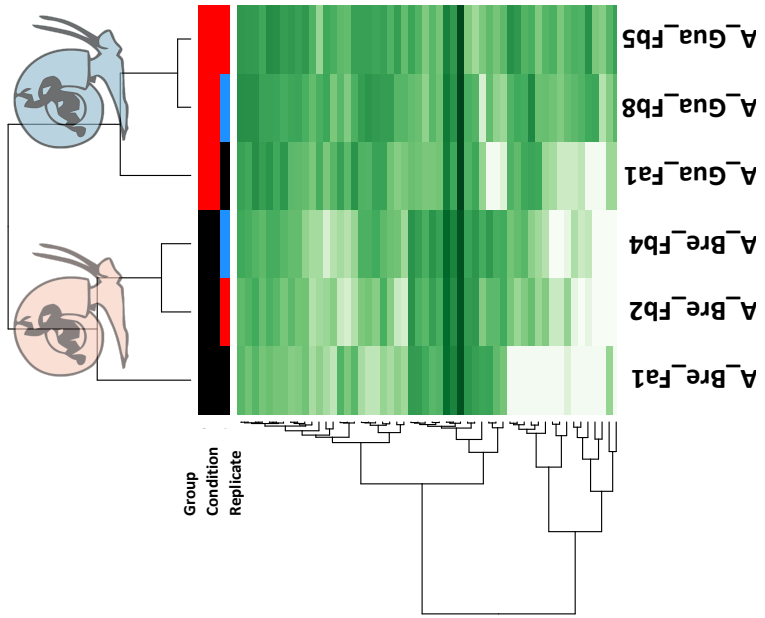
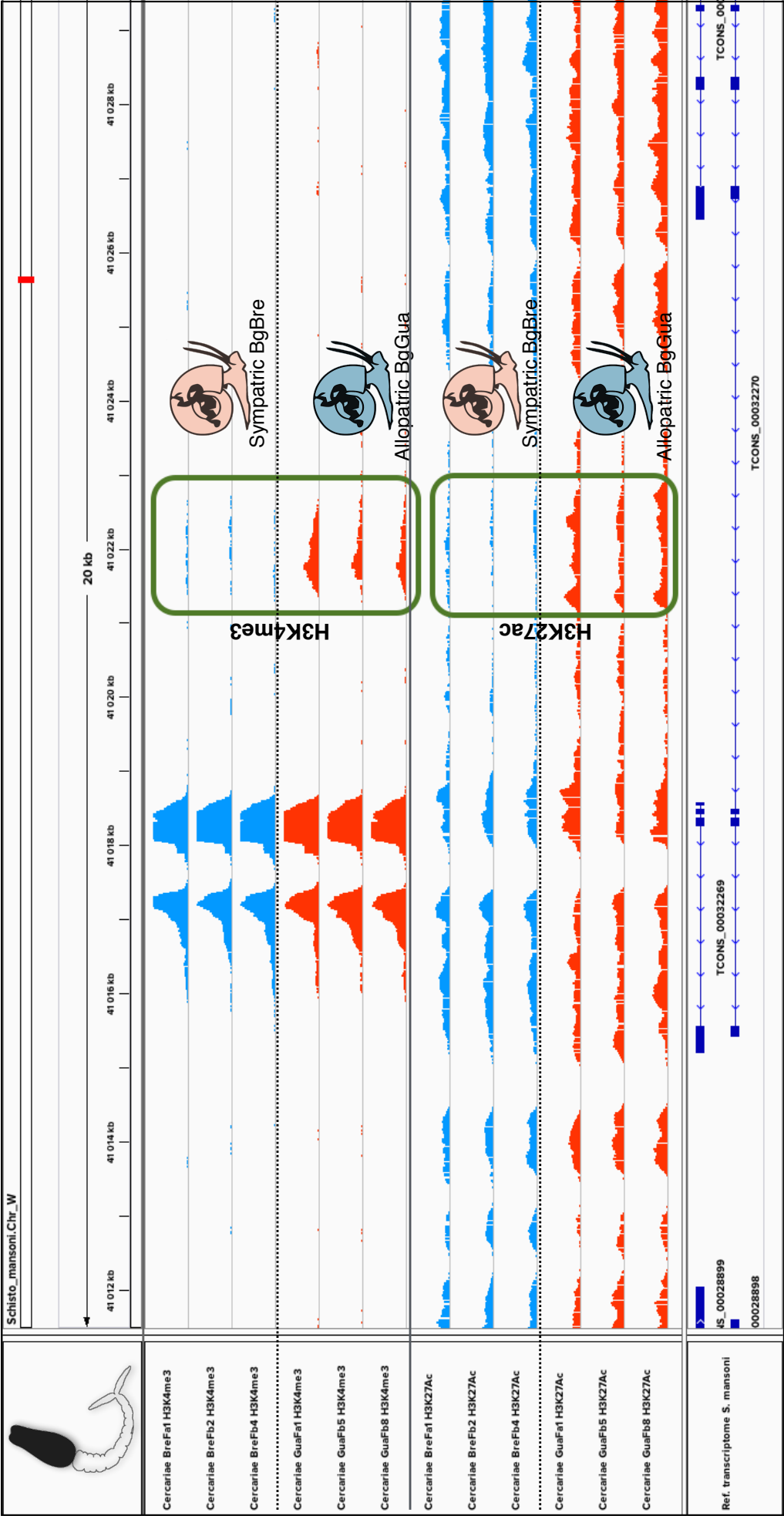
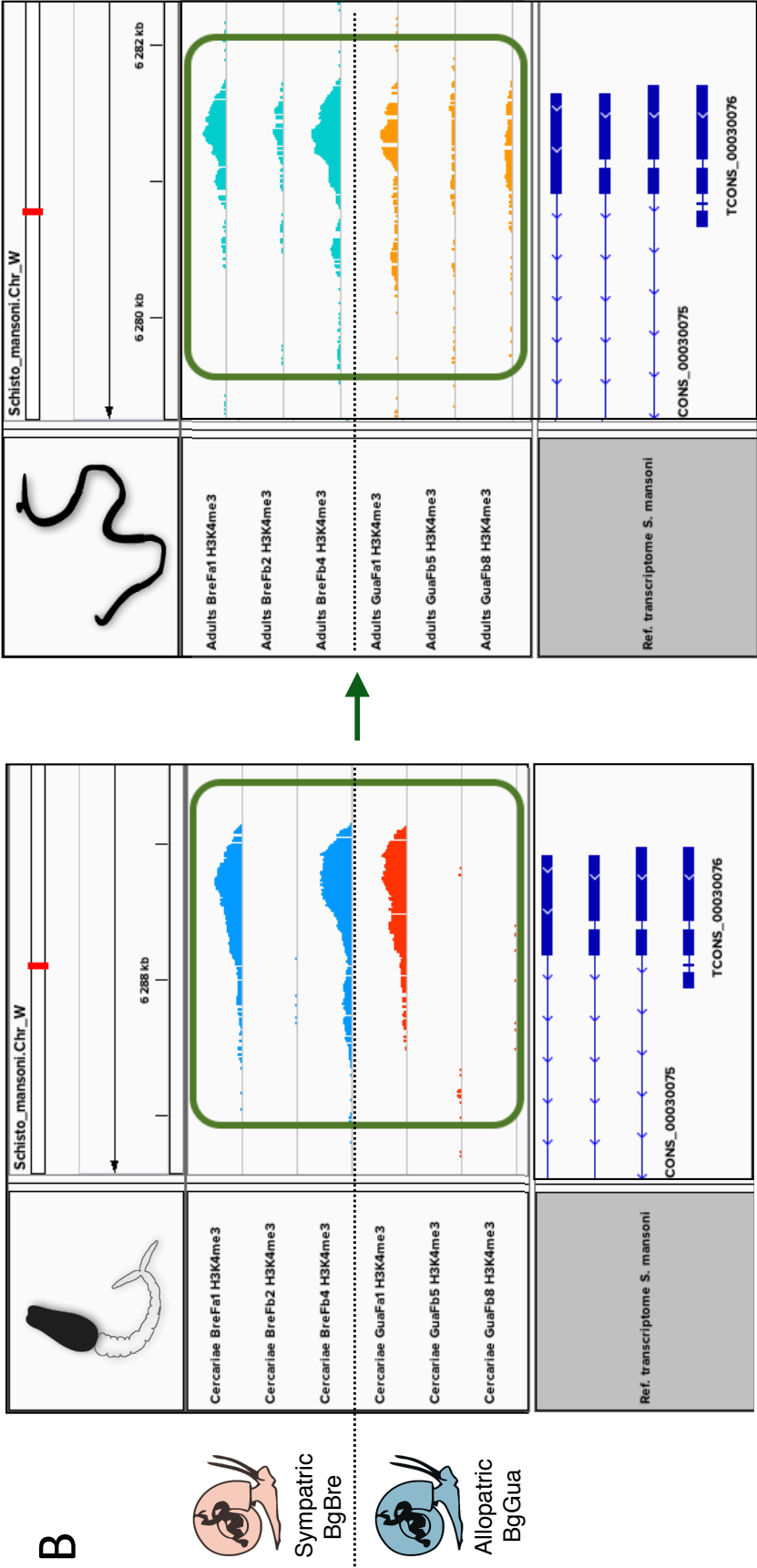
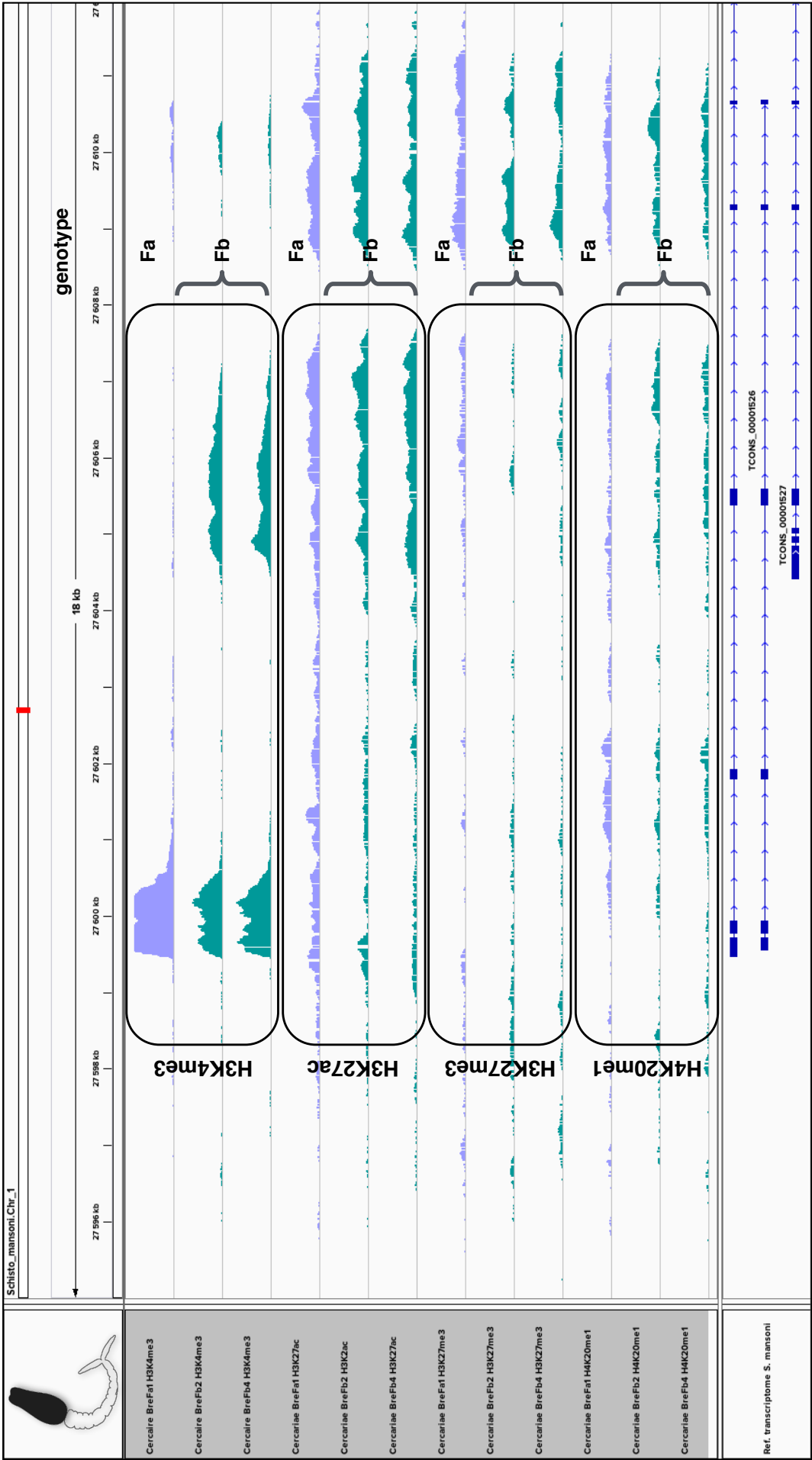
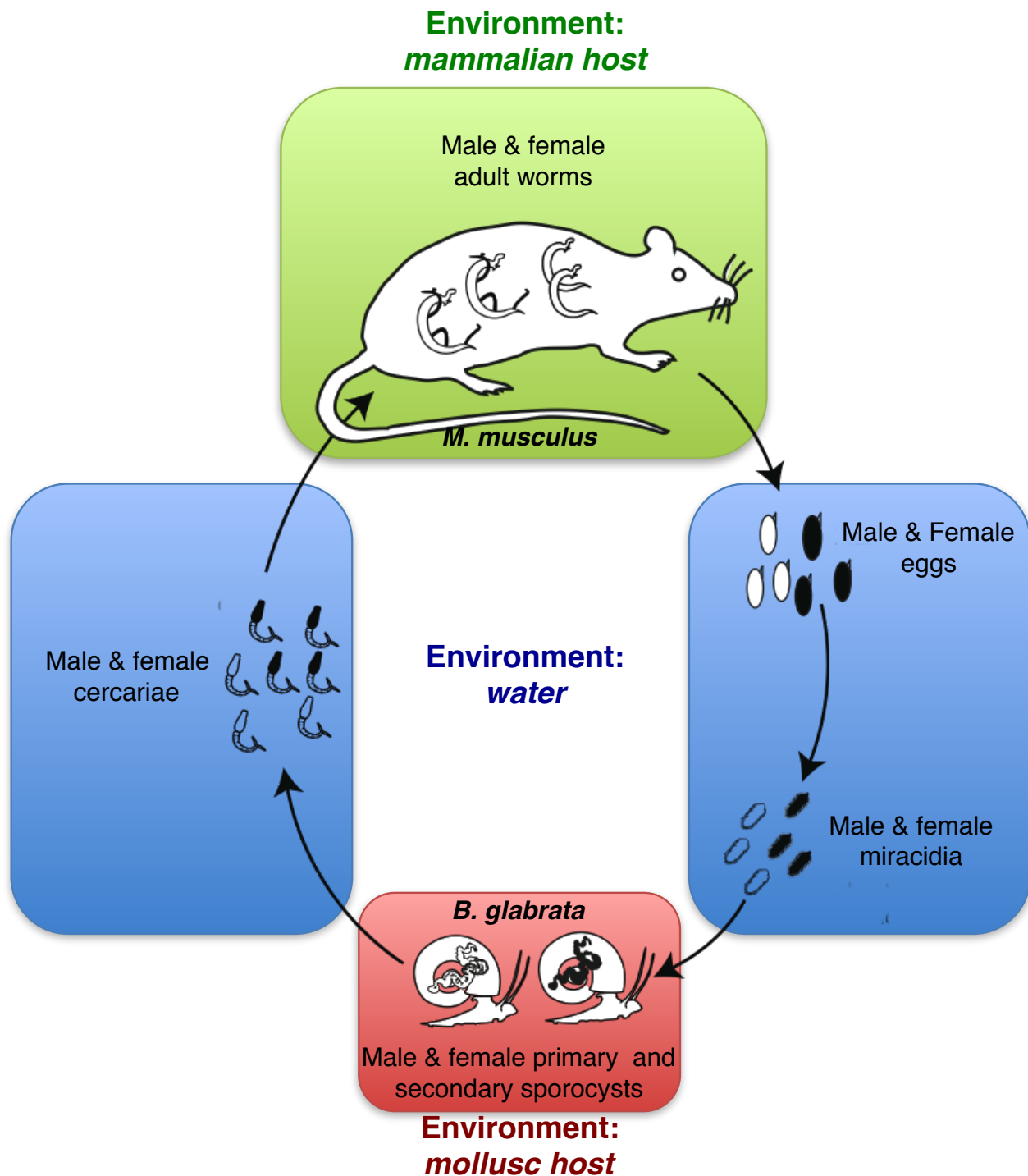


Figure 3

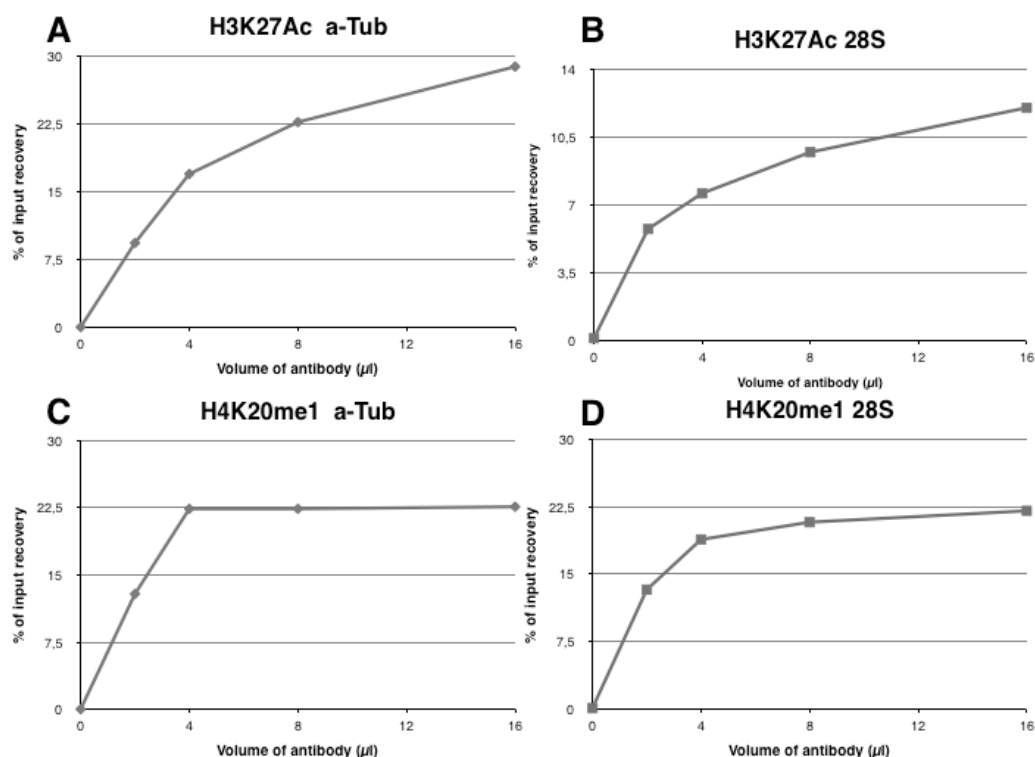




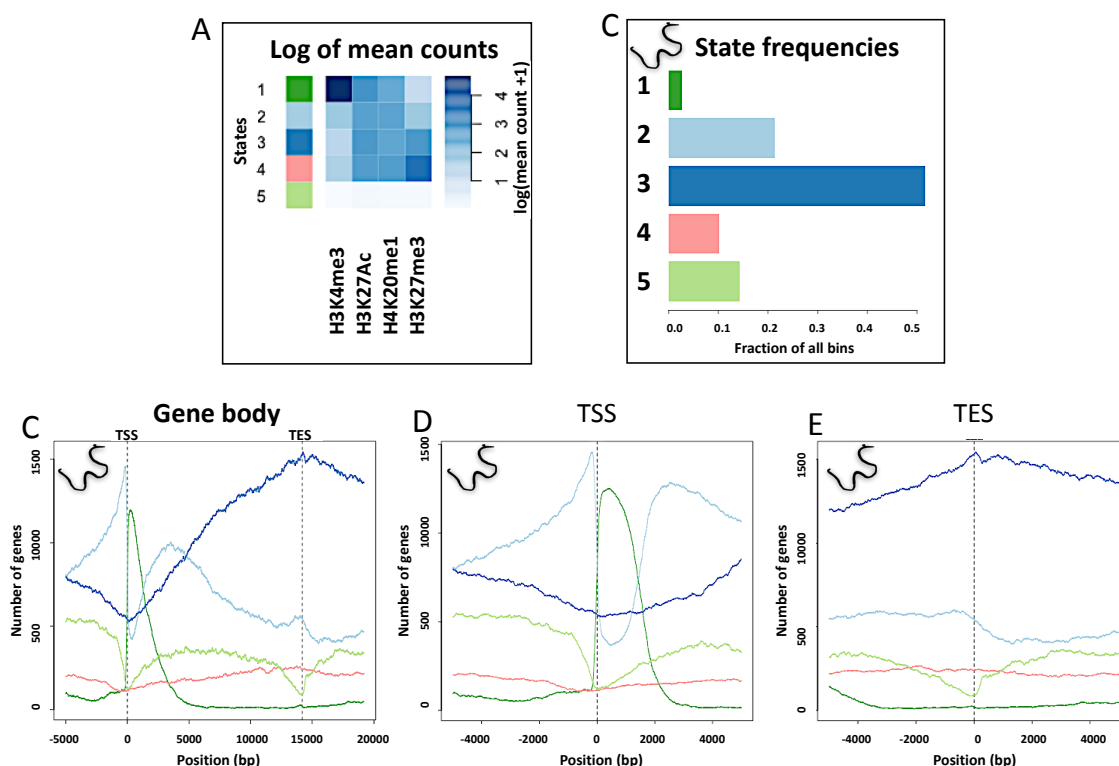




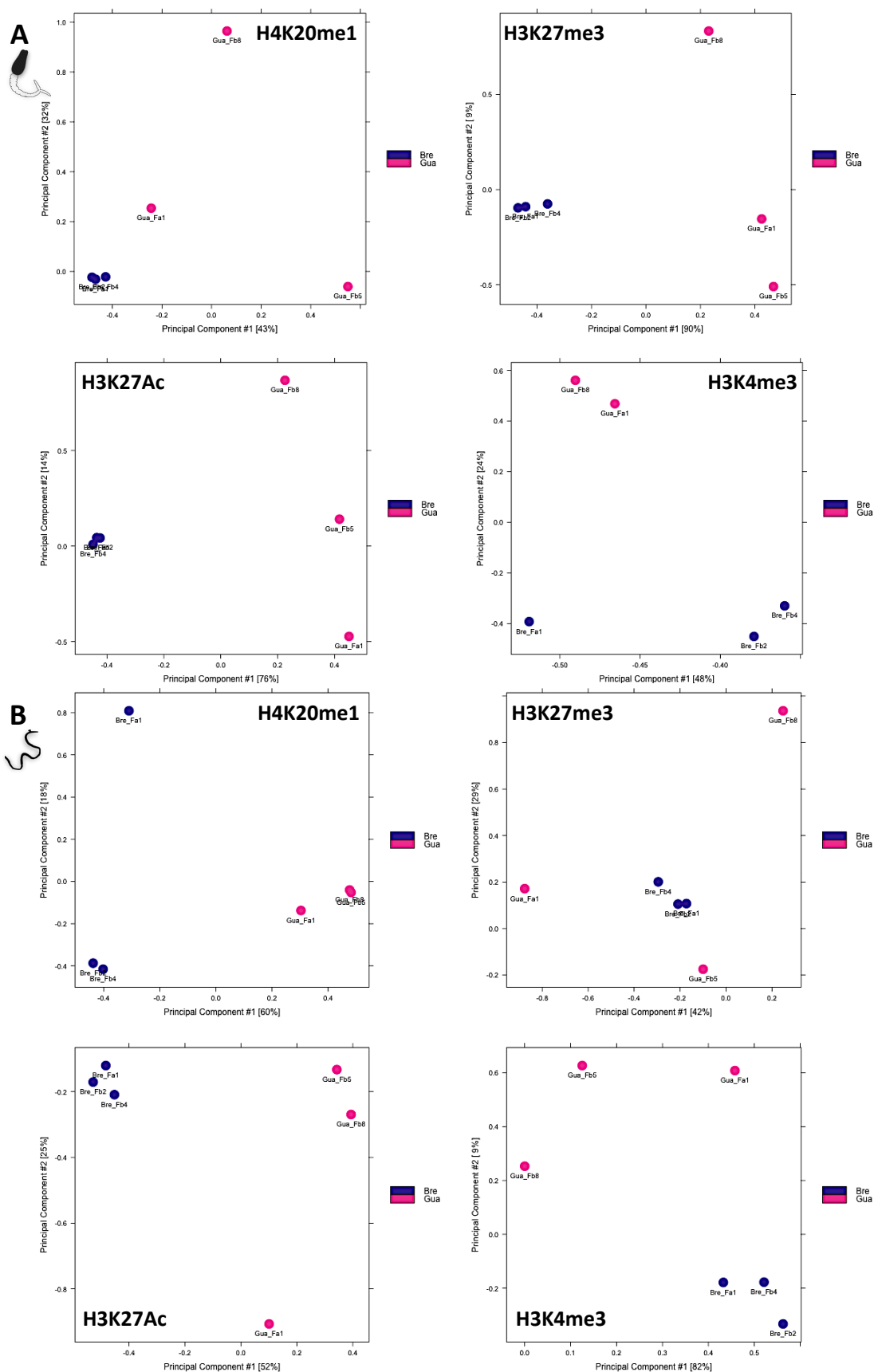
Supplementary Figure 1: Life cycle of the parasite *Schistosoma mansoni*. The endoparasites infect two hosts, (1) a mollusk intermediate host (*Biomphalaria* genus) in which it multiplies asexually as sporocysts and then cercariae, and (2) a vertebrate definitive host (primate or rodent) in which adult worms reproduce sexually. Two types of free-swimming larvae, miracidia and cercariae, are the infective form for the mollusk and vertebrate hosts, respectively. Eggs laid by adult worm couples in the mesenteric vein of the vertebrate host can go through the intestine wall to be excreted in the feces, and hatch in contact with freshwater.



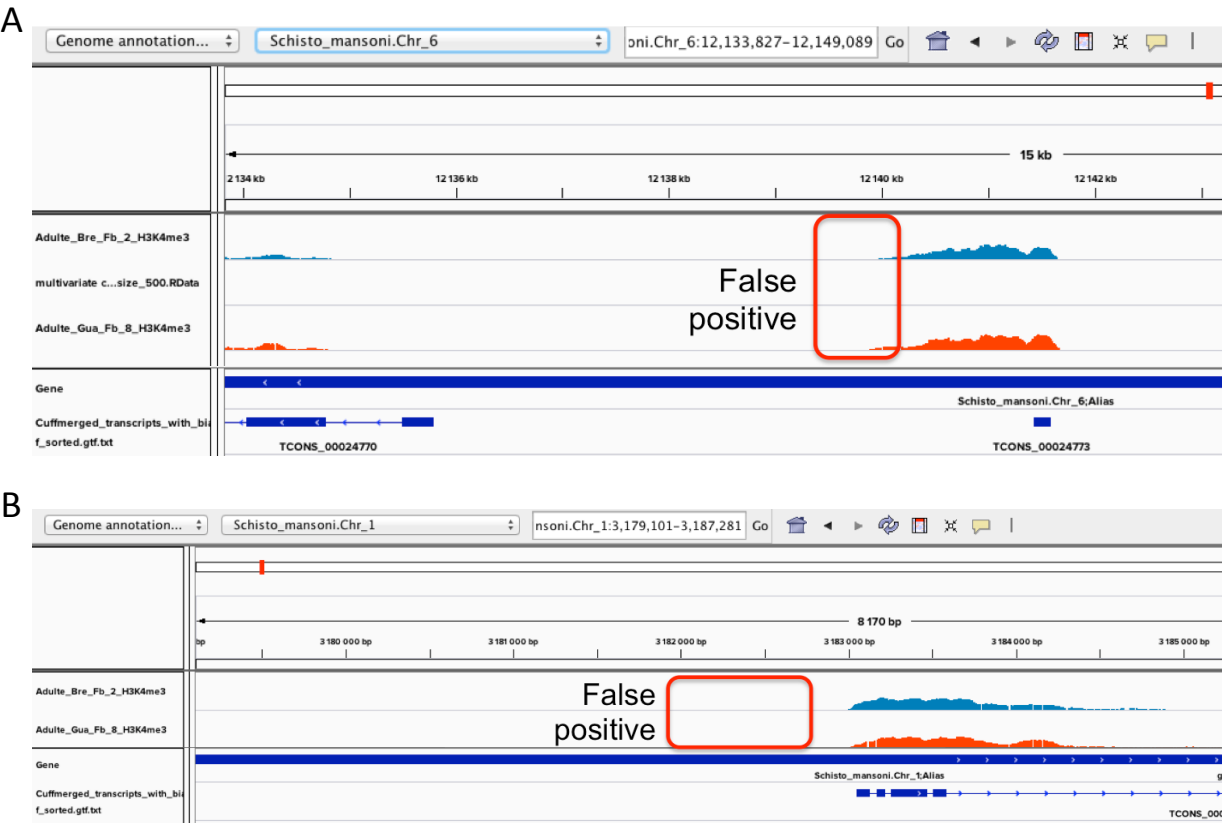
Supplementary Figure 2: ChIP titration, using the same amount of *S. mansoni* chromatin (equivalent to 20 adult worms) and an increasing amount of antibody. Percentage of input recovery of chromatin was measured by qPCR on two housekeeping genes, alpha tubulin and 28S, as described in (Cosseau *et al.* 2009). Titrations were done to measure the efficiency of immunoprecipitation for antibody H3K27ac Abcam cat# ab4729 lot# GR-150-367-2 (**A** and **B**) and H4K20me1 Abcam, cat# ab9051 lot# GR158874-1 (**B** and **C**).



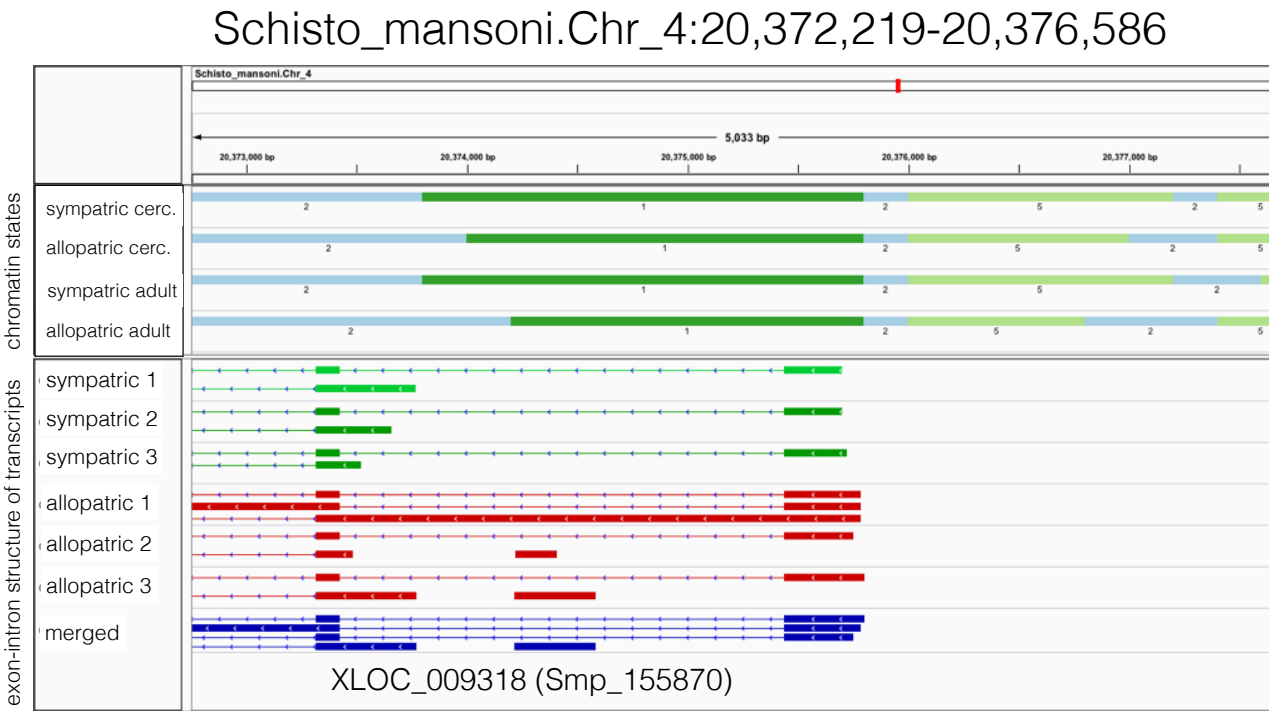
Supplementary Figure 3: Chromatin states identified by EpiCseg in adults. (**A**) Composition in histone marks of the states. Color code of the states are used throughout of the figure. (**B**) Proportion of the genome covered by each state in adults. Enrichment for each state around gene bodies (**C**), transcription start site (TSS) (**D**) and transcription end site (TES) (**E**) of adults. Enrichment profiles were done on 2,428 genes longer than 5 kb.



Supplementary Figure 4: Principal Component Analysis (PCA) for the four studied histone marks in cercariae (**A**) and adults (**B**). Only sites with statistically significant differences (epimutations) were used. Blue circles (Bre) represent sympatric cercariae and adults (*SmBre* in *BgBre*), magenta (Gua) the allopatric conditions (*SmBre* in *BgGua*). Cercariae sample GuaFb5 for H3K4me3 was excluded from analysis (see methods).



Supplementary Figure 5: Examples of two regions (**A** and **B**) that were identified as epimutations (displayed by the red boxes) by DiffBind but did not pass the visual inspection.



Supplementary Figure 6: Example of a region that was not identified as epimutated by DiffBind but shows differences based on EpiCseg chromatin “colors”, judged by visual inspection. The associated gene (putative mitochondrial import inner membrane translocase) is expressed thirteen fold in cercariae from allopatric hosts. Small changes in extent of chromatin states 1 and 5 between sympatric and allopatric hoststs are faithfully transmitted from cercariae to adults

Supplementary File 3: PeakRanger results (number of peaks)

(A) Adults

	Bre_Fa_1	Bre_Fb_2	Bre_Fb_4	Gua_Fa_1	Gua_Fb_5	Gua_Fb_8	average	SD
H3K4me3	7 946	7 928	8 400	7 983	8 946	9 062	8 378	517
H3K27ac	14 517	14 161	8 796	13 095	39 646	4 983	15 866	12 217
H3K27me3	22 106	27 755	21 193	14 830	22 261	16 558	20 784	4 604
H4K20me1	6 600	9 979	8 935	3 725	31 402	6 408	11 175	10 146

(B) Cercariae

	Bre_Fa_1	Bre_Fb_2	Bre_Fb_4	Gua_Fa_1	Gua_Fb_5	Gua_Fb_8	average	SD
H3K4me3	11 726	12 734	12 316	14 827	8 228	11 971	11 967	2 141
H3K27ac	5 358	10 485	2 911	15 039	6 855	14 014	9 110	4 872
H3K27me3	32 205	32 554	18 713	12 616	9 959	28 560	22 435	10 013
H4K20me1	21 262	27 781	14 062	10 756	12 254	18 300	17 403	6 407

Supplementary File 4: (A) Summary

	Environmentally triggered	Genotype specific	Random
Total changes	51	64	64
% Changes in transcribed regions	76,50%	73,40%	79,70%
% Changes at TSS	17,6%	37,5%	37,5%
% Changes at TES	5,9%	12,5%	3,1%
% Changes in Exons	11,8%	1,6%	10,9%
% Changes in Introns	37,3%	14,1%	18,8%
% Changes in promoters (2 kb upstream of TSS)	5,9%	7,8%	9,4%
% Changes in intergenic regions	21,6%	26,6%	20,3%
% of sites including a change in H3K4me3	100%	67,2%	93,80%
% transmitted to adults	0	76,60%	45,3%

Supplementary File 4: (B) List of environmentally dependant epigenetic differences found within cercariae samples

Note: histone combination column describe which mark are different between sympatric (bre) and allopatric (gua) samples at the given position

Position (specific)	Histone combination	Location	Transcribed region?	Gene	Function	Note	Heritability	Enrichment in H3K4me3?	Where are the H3K4me3 peaks?
Schisto_mansoni.Chr.1:35082604-35083988	Bre:H3K4me3	Intron 2 alt	Yes	Snp_141690	hypothetical protein	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.1:38228000-38228500	Bre:H3K4me3	TSS	Yes	Snp_125330	ATP dependent RNA helicase DDX55	Trans-spliced gene	No	Yes	Peak in Bre
Schisto_mansoni.Chr.1:39982500-39983000	Bre:H3K4me3	TSS	Yes	Snp_033130	beta3 glucuronyltransferase	Trans-spliced gene	No	Yes	Peak in Bre
Schisto_mansoni.Chr.1:54177895-54178564	Bre:H3K4me3	TSS	Yes	N/A	N/A	Unannotated transcript	No	Yes	Peak in Bre
Schisto_mansoni.Chr.1:9624744-9625811	Bre:H3K4me3	Intron 3	Yes	Snp_162050	Fas death domain associated protein	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.1_unplaced.SC_0010:4112000-4112500	Bre:H3K4me3	Intron 4	Yes	Snp_201150	Equilibrative nucleoside transporter 1	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.1_unplaced.SC_0076:804500-805000	Bre:H3K4me3	Intron 5	Yes	Snp_176690	rai guanine nucleotide exchange factor with pH	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.1_unplaced.SC_0153:483283-483661	Bre:H3K4me3	Exon 11	Yes	Snp_176000.1	abnormal spindle microcephaly associated	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.2:20253983-20258474	Bre:H3K4me3	Exon 3	Yes	Snp_147260	RAS, family 10, member B like	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.2:21762228-21762788	Bre:H3K4me3	TSS	Yes	Snp_157380	Hypothetical protein	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.2:31839000-31839500	Bre:H3K4me3	Intron 1	Yes	Snp_122790	hypothetical protein	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.2:34366878-34367400	Bre:H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.2:34398500-34399000	Bre:H3K4me3-Gua:H3K27Ac	Intron 8	Yes	Snp_159340	Dynein intermediate chain	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.2_unplaced.SC_0120:266000-267000	Bre:H3K4me3	Exon 7	Yes	Snp_158280	zinc finger protein	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.2_unplaced.SC_0171:484848-486208	Bre:H3K4me3	Promoter	No	Snp_002350	asparagine rich protein	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.3:12057000-12058000	Bre:H3K4me3-Bre:H3K27Ac-Bre:H4K20me1	TSS	Yes	Snp_010110	E3 ubiquitin protein ligase MARCH5	Alternative TSS	No	Yes	Peak in Bre
Schisto_mansoni.Chr.3:19021502-1902221	Bre:H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.3:22302500-22303000	Bre:H3K4me3-Gua:H3K27Ac	Promoter	Yes	Snp_160860	tetratricopeptide repeat protein 8	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.3:8474500-8475500	Bre:H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.3:9246000-9246500	Bre:H3K4me3	Exon 6	Yes	Snp_145980	excitatory amino acid transporter 3	Alternative splicing	No	Yes	Longer peaks in Bre
Schisto_mansoni.Chr.4:12227500-12228000	Bre:H3K4me3	TSS	Yes	Snp_202660	hypothetical protein	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.4:15830850-1583110	Bre:H3K4me3-Bre:H4K20me1	Intron 3	Yes	Snp_153970	RPL protein	Alternative TSS	No	Yes	Peak in Bre
Schisto_mansoni.Chr.4:19299000-19299500	Bre:H3K4me3-Gua:H3K27me3-Gua:H4K20me1	TSS	Yes	Snp_169630	exportin 6	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.4:28484000-28484500	Bre:H3K4me3	Intron 6	Yes	Snp_149220	G_T mismatch specific thymine DNA glycosylase	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.4:4965000-4965500	Bre:H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.4_unplaced.SC_0032:264500-265500	Bre:H3K4me3	Intergenic	No	Snp_016630	TPA: Sm_alphaT1	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.6:16666000-16666500	Bre:H3K4me3-Gua:H3K27Ac	Promoter	Yes	SnrRNA_01223_Gli_TTG.11	trimeric G protein alpha o subunit	Trans-spliced gene	No	Yes	Peak in Bre
Schisto_mansoni.Chr.6:16957500-16958000	Bre:H3K4me3	TSS	Yes	Snp_163960	Wlms tumor protein 1 interacting protein	Alternative TSS	No	Yes	Peak in Bre
Schisto_mansoni.Chr.6:18806500-18807000	Bre:H3K4me3	Intron 1	Yes	Snp_077560	Hippocampus abundant transcript 1 protein	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.6:18994000-18994500	Bre:H3K4me3-Gua:H4K20me1	Intron 8	Yes	Snp_123680	hypothetical protein	Alternative TSS	No	Yes	Peak in Bre
Schisto_mansoni.Chr.6:2750000-2750500	Bre:H3K4me3-Gua:H3K27Ac	Intron 3	Yes	Snp_141980	cAMP specific 3'5' cyclic phosphodiesterase 4D	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.6:8851651-8852668	Bre:H3K4me3	Intron 7	Yes	Snp_141980	cAMP specific 3'5' cyclic phosphodiesterase 4D	Trans-spliced gene	No	Yes	Peak in Bre
Schisto_mansoni.Chr.6:8851954-8852606	Bre:H3K4me3-Gua:H3K27Ac-Gua:H3K27me3	Intron 9	Yes	Snp_141980	cAMP specific 3'5' cyclic phosphodiesterase 4D	Trans-spliced gene	No	Yes	Peak in Bre
Schisto_mansoni.Chr.7:3322000-3323000	Bre:H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.7:18929518-18930078	Bre:H3K4me3	Intron 1 / Ex	Yes	N/A	N/A	Alternative splicing	No	Yes	Peak in Bre
Schisto_mansoni.Chr.7:21685500-21689500	Bre:H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.7:26688097-26688880	Bre:H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.7:32870500-32871000	Bre:H3K4me3-Gua:H3K27Ac	TSS	Yes	N/A	N/A	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.7:37486844-37488816	Bre:H3K4me3	Intron 4	Yes	Snp_146910	choline acetyltransferase	Trans-spliced gene	No	Yes	Peak in Bre
Schisto_mansoni.Chr.7:40215600-4022000	Gua:H3K4me3-Gua:H3K27Ac	Intron 10	Yes	Snp_151290	multidrug and toxin extrusion protein 2	N/A	No	Yes	Peak in Gua
Schisto_mansoni.Chr.7:51973673-51974617	Bre:H3K4me3	Intron 7	Yes	Snp_039820	cysteine protease atg4b	N/A	No	Yes	Peak in Bre
Schisto_mansoni.Chr.7:53980074-53980799	Bre:H3K4me3	Exon 1	Yes	N/A	N/A	Unannotated transcript	No	Yes	Peak in Bre
Schisto_mansoni.SC_0013:1787452-1788526	Bre:H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre
Schisto_mansoni.SC_00371703601-1704948	Bre:H3K4me3	Intron 1	Yes	Snp_130050	PDZ domain containing RING finger protein	N/A	No	Yes	Peak in Bre
Schisto_mansoni.SC_0037158500-59000	Bre:H3K4me3	TSS	Yes	N/A	TPA: Pol Polyproteine	N/A	No	Yes	Peak in Bre
Schisto_mansoni.SC_0049:463408-464854	Bre:H3K4me3	TSS	Yes	Snp_125070	cystathionine beta synthase	Alternative TSS	No	Yes	Peak in Bre
Schisto_mansoni.SC_0080:606880-607457	Bre:H3K4me3	Intron 8	Yes	Snp_133370	tetratricopeptide repeat protein 39B	N/A	No	Yes	Peak in Bre
Schisto_mansoni.SC_0178:546500-547000	Bre:H3K4me3-Gua:H3K27Ac	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre
Schisto_mansoni.SC_0179:272000-272500	Bre:H3K4me3	TSS	Yes	Snp_076620	RH1 Inducible coiled coil protein 1	Alternative TSS	No	Yes	Peak in Bre
Schisto_mansoni.SC_0219:229000-229500	Bre:H3K4me3-Gua:H3K27Ac	Exons 12-14	Yes	Snp_098760	F box:WD repeat containing protein	N/A	No	Yes	Peak in Bre

Supplementary File 4: (C) List of spontaneous epigenetic differences found between cercariae samples

Note: histone combination column describe which mark are different between the two sets of samples (arbitrarily named R1 and R2) at a given position

Position (specific)	Histone combination	Location	Transcribed region?	Gene	Function	Note	Heritability	Enrichment in H3K4me3?	Where are the H3K4me3 peaks?
Schisto_mansoni.Chr.16324000-6324500	R1H3K4me3	TSS	Yes	N/A	PolyA stretch	PolyA stretch (with transcripts)	Partial (not all samples)	Yes	Peak in Bre Fat & Fb2, Gua Fb5 & Fb8
Schisto_mansoni.Chr.12146500-2146900	R1H3K4me3,R1H4K20me1	TSS	Yes	N/A	hypothetical protein	N/A	No	Yes	Peak in Bre F2 Bre
Schisto_mansoni.Chr.139897043-39897539	R1H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre Fat & Fb4, Gua Fat & Fb5 & Fb8
Schisto_mansoni.Chr.142745000-42745500	R1H3K4me3	Intron1	Yes	Smp_144600	guanine nucleotide exchange factor	N/A	No	Yes	Peak in Bre Fat & Bre Fb2
Schisto_mansoni.Chr.15059863-50598656	R1H3K4me3,R1H3K27Ac	TSS	Yes	Smp_148700	leukocyte receptor cluster	Trans-spliced gene, peak size variations	YES	Yes	Peak in Bre Fat & Fb4, Gua Fat & Fb8
Schisto_mansoni.Chr.159029500-59029500	R1H3K4me3	Intergenic	No	N/A	N/A	Peak size variations	YES	Yes	Peak in Bre Fat (small) & Fb2 & Fb4 (small), Gua Fat (small) Fb5 & Fb 8
Schisto_mansoni.Chr.165343500-65343500	R1H3K4me3	Intron 2 alt	Yes	Smp_000755	family M3 non peptidase ue (M3 family)	Partial (not all samples)	Partial	Yes	Peak in Gua Fb5
Schisto_mansoni.Chr.19347757-9348822	R1H3K4me3,R1H3K27Ac	Intron 2	Yes	Smp_197070	kinesin protein KIF6	N/A	YES	Yes	Peak in Bre Fb2 & Gua Fb8
Schisto_mansoni.Chr.19348000-9349000	R1H3K4me3	TSS	Yes	N/A	N/A	N/A	YES	Yes	Peak in Bre Fb2 & Gua Fb8
Schisto_mansoni.Chr.Lupluced SC_0071618500-16190	R1H3K4me3	Intron 1	Yes	N/A	GTP- protein coupled receptor fragment, putative	N/A	YES	Yes	Peak in Bre Fat & Fb4, Gua Fat & Fb8
Schisto_mansoni.Chr.Lupluced SC_0071486000-48650	R1H3K4me3,R2H3K27me3	TSS	Yes	N/A	TPA: endonuclease-reverse transcriptase	Peak size variations	YES	Yes	No peak in Bre Fb2
Schisto_mansoni.Chr.212442000-12443000	R1H3K4me3,R2H3K27me3	Intergenic	No	N/A	N/A	Peak size variations	YES	Yes	No peak in Bre Fb2
Schisto_mansoni.Chr.213379000-13379500	R1H3K4me3	Exon 2	Yes	Smp_148760	unconventional profilin rpb5 interactor	N/A	YES	Yes	Peak in Bre Fb4, Gua Fat & Fb5
Schisto_mansoni.Chr.219862500-19863500	R1H3K4me3	TSS	Yes	Smp_147320	camp dependent protein kinase regulatory ch	N/A	No	Yes	Peak in Bre Fb4 & Gua Fat
Schisto_mansoni.Chr.22145000-2145500	R1H3K4me3	TSS	Yes	Smp_165390	hypothetical protein	Peak size variations	YES	Yes	Peak in Bre Fat & Fb2 (small) & Fb4, Gua Fat (small) & Fb5 & Fb8 (small)
Schisto_mansoni.Chr.227740500-27741000	R1H3K4me3	Promoter	Yes	Smp_146010	insulysin unit 3 (Mf6 family)	Peak size variations	YES	Yes	Different Peak Shape in Fat Bre & Gua
Schisto_mansoni.Chr.26523500-6524500	R1H3K4me3	Promoter	Yes	N/A	TPA: Reverse transcriptase SR3-Right	N/A	No	Yes	Peak in Bre Fat & Bre Fb4
Schisto_mansoni.Chr.285000-85500	R1H3K4me3	Intron 4	Yes	Smp_212600	protein kinase	N/A	No	Yes	Peak in Bre Fat & Bre Fb4
Schisto_mansoni.Chr.313083000-13085000	R1H3K4me3	Exon1	Yes	N/A	N/A	Also correspond to Intron 2 of Smp_129110	No	Yes	Peak in Bre Fb4
Schisto_mansoni.Chr.327093000-27030000	R1H3K4me3	TSS	Yes	Smp_015050	cholesterol/ethanolamine kinase	Peak size variations	No	Yes	Different peak shapes between all samples
Schisto_mansoni.Chr.371652964-71653387	R1H3K4me3	TSS	Yes	N/A	N/A	Peak size variations	No	Yes	Peak in Bre Fat & Fb2, Gua Fat & Fb8
Schisto_mansoni.Chr.3uplaced SC_00441574000-57450	R1H3K4me3	Promoter	Yes	N/A	N/A	Peak size variations	No	Yes	No peak in Gua Fb8
Schisto_mansoni.Chr.41965000-14965500	R1H3K4me3	Intron 2	Yes	Smp_058980	oxysterol binding protein 1	Alternative splicing	YES	Yes	Peak in Bre Fb2
Schisto_mansoni.Chr.41964500-19645500	R1H3K4me3	Promoter	Yes	Smp_160680	breast cancer type 2 susceptibility protein	N/A	YES	Yes	Peak in Bre Fb2, Gua Fb5 & Fb8
Schisto_mansoni.Chr.419520260-1952352	R1H3K4me3	Intron2	Yes	Smp_160190	primase, DNA, polypeptide 1 (9 ID)a	Peak size variations	No	Yes	Different peak shape in Bre Fat & Fb4, Gua Fat
Schisto_mansoni.Chr.41924869000-424869500	R1H3K4me3	TSS	Yes	Smp_163640	nuclear pore complex protein Nup54	Alternative TSS	No	Yes	Peak in Gua Fb8
Schisto_mansoni.Chr.612933000-129340000	R1H3K4me3	Promoter	Yes	Smp_055050	40S ribosomal protein S17	Peak size variations	YES	Yes	Different peak shape in Bre Fb4
Schisto_mansoni.Chr.61533500-1534000	R1H3K4me3	TSS	Yes	N/A	N/A	N/A	No	Yes	Peak in Bre Fb2
Schisto_mansoni.Chr.616992000-16983000	R1H3K4me3,R2H3K27me3	Intergenic	No	N/A	N/A	N/A	YES	Yes	Bre Fb4, Gua Fb5 & Fb8
Schisto_mansoni.Chr.65288500-5289000	R1H3K4me3	Intron 3	Yes	Smp_213260	hypothetical protein	Trans-spliced gene	YES	Yes	Peak in Gua Fb8
Schisto_mansoni.Chr.71612000-1612500	R1H3K4me3	Intron 4	Yes	Smp_164040	hypothetical protein	N/A	No	No	N/A
Schisto_mansoni.Chr.7171500-172500	R1H3K4me3	Intergenic	No	Smp_127770	ras GTP exchange factor	N/A	No	Yes	Peak in Bre & Gua Fat
Schisto_mansoni.Chr.73322000-3322500	R1H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Bre Fat & Fb2
Schisto_mansoni.Chr.756500-57500	R1H3K4me3	Intergenic	No	N/A	N/A	N/A	Partial (not all samples)	Yes	Peak in Bre Fat & Bre Fb2
Schisto_mansoni.Chr.7806000-806500	R1H3K4me3	Exon 2	Yes	Smp_097370	fructose 1,6 biphosphatase	N/A	No	Yes	Peak in Bre, Gua Fb8
Schisto_mansoni.Chr.79699500-9700000	R1H3K4me3	Exon 2	Yes	Smp_139420	hypothetical protein	N/A	No	Yes	Peak in Bre Fb2
Schisto_mansoni.Chr.W7339620-7340523	R1H3K4me3	TSS	Yes	N/A	TPA:Reverse transcriptase	Peak size variations	YES	Yes	Peak in Bre Fb2 & Gua Fb 8
Schisto_mansoni.Chr.W9549500-9548500	R1H3K4me3	TSS	No	N/A	TPA: endonuclease-reverse transcriptase	Sm Perere transposable element	PARTIAL	Yes	Peak in Bre Fat & Bre Fb2
Schisto_mansoni.Chr.W12566500-12567000	R1H3K4me3,R1H3K27Ac	TSS	Yes	N/A	N/A	N/A	No	Yes	Peak in Bre Fat & Bre Fb2
Schisto_mansoni.Chr.W14603000-14605500	R1H3K4me3,R1H3K27me3	Intron 1	Yes	Smp_096390	16 kda calcium binding protein	N/A	No	Yes	Peak in Bre Fb4 & Gua Fat
Schisto_mansoni.Chr.W18705500-18707000	R1H3K4me3	Intergenic	No	N/A	N/A	N/A	No	Yes	Peak in Bre Fb4
Schisto_mansoni.Chr.W18926000-18927500	R1H3K4me3	TSS	Yes	N/A	N/A	N/A	YES	Yes	Peak in Bre Fb4
Schisto_mansoni.Chr.W19926000-19926500	R1H3K4me3	TSS	Yes	Smp_148770	hypothetical protein	N/A	No	Yes	Peak in Bre Fat & Fb4, Gua Fat & Fb8
Schisto_mansoni.Chr.W23768500-23769000	R1H3K4me3	TSS	Yes	N/A	N/A	N/A	No	Yes	Peak in Bre Fat & Bre Fb2
Schisto_mansoni.Chr.W230380500-30381500	R1H3K4me3,R1H3K27Ac	Promoter	Yes	Smp_155380	hypothetical protein	N/A	No	Yes	Peak in Bre Fat & Fb4, Gua Fat & Fb8
Schisto_mansoni.Chr.W3663500-3664000	R1H3K27Ac,R2H4K20me1	TSS	Yes	Smp_1695801	hypothetical protein	N/A	No	Yes	N/A
Schisto_mansoni.Chr.W3727087-37271915	R1H3K4me3,R1H3K27Ac,R2TSS	TSS	Yes	N/A	N/A	Peak size variations	YES	Yes	Peak in Bre Fb2 (small) & Fb4, Gua Fb8 (small)
Schisto_mansoni.Chr.W40322000-40322500	R1H3K4me3	TSS	Yes	N/A	T repeats	N/A	YES	Yes	Peak in Bre Fat, Gua Fat & Fb5
Schisto_mansoni.Chr.W51532500-51534000	R1H3K4me3	Intron 1	Yes	Smp_139680	retinol dehydrogenase 11	N/A	YES	Yes	Peak in Bre Fb4 & Gua Fat
Schisto_mansoni.Chr.W58587063-58587546	R1H3K4me3,R2H3K27me3	Intron 1	Yes	N/A	N/A	N/A	YES	Yes	Peak in Bre Fat & Fb4, Gua Fat & Fb8
Schisto_mansoni.Chr.W5900000-59005000	R1H3K4me3	Intergenic	No	N/A	N/A	Peak size variations	No	Yes	Bre Fb2 & Gua Fb5
Schisto_mansoni.Chr.W59259000-59300000	R1H3K4me3	TSS	Yes	N/A	N/A	Peak size variations	No	Yes	Peak in Bre Fat & Fb2, Gua Fat & Fb8
Schisto_mansoni.Chr.W62876786289120	R1H3K4me3	TSS	Yes	N/A	TPA:Reverse transcriptase	N/A	YES	Yes	Peak in Bre Fat & Fb8, Gua Fat
Schisto_mansoni.Chr.W9549500-9545000	R1H3K4me3	Intergenic	No	N/A	N/A	N/A	Partial	Yes	Bre Fat & Fb2
Schisto_mansoni.SC_001561500-670732	R1H3K4me3	Intergenic	No	N/A	N/A	N/A	YES	Yes	Peak in Bre Fb2, Gua Fb5 & Fb8
Schisto_mansoni.SC_0028065000-650500	R1H3K4me3	TSS	Yes	N/A	N/A	N/A	YES	Yes	Peak in Gua Fb8
Schisto_mansoni.SC_0129262500-63000	R1H3K4me3	Intergenic	Yes	N/A	N/A	N/A	No	Yes	Peak in Bre Fb2, Gua Fat & Fb8
Schisto_mansoni.SC_0136512000-572500	R1H3K27Ac,R2H4K20me1	TSS	Yes	Smp_158050	Trans-spliced gene	N/A	No	N/A	N/A
Schisto_mansoni.SC_0211240500-242000	R1H3K4me3	Exon 4	Yes	Smp_199120	protein FAM114A2	protein FAM114A2	YES	Yes	Peak in Gua Fb8
Schisto_mansoni.SC_0212318500-319000	R1H3K4me3,R2H4K20me1	Intergenic	No	N/A	(TT AGCG) repeat	Peak size variations	No	Yes	Peak in Gua Fb5
Schisto_mansoni.SC_0226121000-121500	R1H3K4me3	Exon 7	Yes	Smp_175760	dna) subfamily c	N/A	No	Yes	Peak in Gua Fb8
Schisto_mansoni.SC_21781500-2000	R1H3K4me3	TSS	Yes	N/A	N/A	N/A	Partial (not all samples)	Yes	Peak in Gua Fb5
Schisto_mansoni.SC_25270-1000	R1H4K20me1	Exon1	Yes	N/A	N/A	N/A	No	N/A	N/A

Supplementary File 4: (D) List of epigenetic differences between cercariae genotype Fa & Fb (both Fb2 & Fb4)

Note: histone combination column describe which mark are different between cercariae genotype Fa and Fb at the given position

Position (specific)	Histone combination	Location	Transcribed region?	Gene	Function	Note	Heritability	Enrichment in H3K4me3?	Where are the H3K4me3 peaks?
Schisto_mansonl.Chr.1.15545400-15555000	Fa H3K4me3-Fb H3K27Ac	Intergenic	No	N/A	N/A	N/A	Partial (not all marks)	Yes	Peak in Fa
Schisto_mansonl.Chr.1.17467000-17467500	Fa H3K4me3-Fa H3K27Ac	TSS	Yes	N/A	N/A	N/A	Yes	Yes	Peak in Fa
Schisto_mansonl.Chr.1.27606500-27607000	Fa H3K27me3-Fb H3K4me3-Fb H4K20me1	TSS	Yes	Smp_196620	glycerophosphodiester phosphodiesterase	Trans-spliced gene	Yes	Yes	Peak in Fa
Schisto_mansonl.Chr.1.32544500-32544500	Fa H3K27me3-Fb H4K20me1	Promoter	Yes	N/A	N/A	Strong difference	Yes	No	Peak in Fb
Schisto_mansonl.Chr.1.38228000-38228500	Fa H3K4me3-Fa H4K20me1-Fb H3K27me3	TSS	Yes	Smp_21810	centromere/kinetochore protein zwf0	Trans-spliced gene	No	Yes	N/A
Schisto_mansonl.Chr.1.39512000-39513000	Fa H3K4me3-Fa H3K27Ac	TSS	Yes	N/A	N/A	N/A	No	Yes	N/A
Schisto_mansonl.Chr.1.39878600-39889700	Fa H3K27Ac-Fa H3K27me3	Intergenic	No	N/A	N/A	N/A	No	No	N/A
Schisto_mansonl.Chr.1.39887500-39889500	Fa H3K4me3-Fa H3K27Ac-Fa H3K27me3	Promoter	Yes	Smp_200570	hypothetical protein	N/A	Partial (not all marks)	Yes	Stronger peak in Fb
Schisto_mansonl.Chr.1.58049000-58049000	Fb H3K27Ac-Fb H3K27me3-Fb H4K20me1	Exons 10-12	Yes	Smp_157590	Protein VTS1	Strong difference (deletion)	Yes	No	N/A
Schisto_mansonl.Chr.1.50948690-50949390	Fb H3K4me3	Intergenic	No	N/A	N/A	N/A	Yes	No	Peak in Fb
Schisto_mansonl.Chr.1.unplaced.SC_0010632500-63300	Fa H3K4me3-Fa H3K27Ac	TSS	Yes	Smp_175680	survival motor neuron protein 1	Trans-spliced gene	No	Yes	N/A
Schisto_mansonl.Chr.1.unplaced.SC_0076115000-116500	Fa H3K4me3	Intergenic	No	N/A	N/A	N/A	Yes	Yes	Peak in Fa
Schisto_mansonl.Chr.2.12424000-12444000	Fa H3K27Ac-Fa H3K27me3-Fa H4K20me1-Fb H3K4me3	Intergenic	No	N/A	N/A	N/A	Yes	Yes	Peak in Fa
Schisto_mansonl.Chr.2.13368500-13369500	Fa H3K4me3-Fb H3K4me3	TSS	Yes	Smp_148250	protein unc 79	N/A	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.2.21479500-21481500	Fa H3K4me3-Fa H3K27Ac-Fa H4K20me1-Fb H3K27me3	TSS	Yes	Smp_148250	protein unc 79	N/A	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.2.22552000-22552500	Fa H3K4me3-Fa H3K27Ac-Fb H3K27me3	Promoter	Yes	N/A	N/A	N/A	Partial (not all marks)	Yes	Peak in Fa
Schisto_mansonl.Chr.2.27440000-27440500	Fa H3K4me3-Fb H3K27Ac-Fb H3K4me3-Fb H3K27me3	Promoter	Yes	Smp_146010	insulin unit 3 (M16 family)	Peak size variations	No	Yes	Peak in Fa
Schisto_mansonl.Chr.3.26409500-6401000	Fa H3K27me3-Fb H4K20me1	TSS	Yes	N/A	N/A	Peak size variations	Partial (not all marks)	Yes	Taller peaks in Fa
Schisto_mansonl.Chr.3.27790007-732300	Fb H3K27Ac-Fb H3K27me3-Fb H4K20me1	Intergenic	No	N/A	N/A	Whole region with different chromatin profile	Yes	No	N/A
Schisto_mansonl.Chr.3.27730000-731500	Fb H3K27Ac-Fb H3K27me3-Fb H4K20me1	Intron 1	Yes	Smp_032690	Hypothetical protein	Whole region with different chromatin profile	Yes	No	N/A
Schisto_mansonl.Chr.3.27320000-733000	Fb H3K4me3-Fb H3K27Ac-Fb H3K27me3-Fb H4K20me1	Intron 1	Yes	Smp_032690	Hypothetical protein	Whole region with different chromatin profile	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.3.27410000-745000	Fb H3K27me3-Fb H4K20me1	TSS	Yes	Smp_032690	Hypothetical protein	Whole region with different chromatin profile	Yes	No	N/A
Schisto_mansonl.Chr.3.2748500-756000	Fb H3K4me3-Fb H3K27Ac-Fb H3K27me3-Fb H4K20me1	TSS	Yes	Smp_032710	Hypothetical protein	Whole region with different chromatin profile	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.3.2768000-770500	Fb H3K27me3-H4K20me1	Intergenic	No	N/A	N/A	Whole region with different chromatin profile	Yes	No	N/A
Schisto_mansonl.Chr.3.2773500-778000	Fb H3K27Ac-Fb H3K27me3-Fb H4K20me1	Intergenic	No	N/A	N/A	Whole region with different chromatin profile	Yes	No	N/A
Schisto_mansonl.Chr.3.2788500-795000	Fb H3K4me3-Fb H3K27Ac-Fb H3K27me3-Fb H4K20me1	Intergenic	No	N/A	N/A	Whole region with different chromatin profile	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.3.2812500-813500	Fb H3K27Ac-Fb H3K27me3-Fb H4K20me1	Intergenic	No	N/A	N/A	Whole region with different chromatin profile	Yes	No	N/A
Schisto_mansonl.Chr.3.2812500-817000	Fb H3K27Ac-Fb H3K27me3-Fb H4K20me1	Intergenic	No	N/A	N/A	Whole region with different chromatin profile	Yes	No	N/A
Schisto_mansonl.Chr.3.unplaced.SC_033216500-17500	Fb H3K4me3-Fb H3K27Ac	Intergenic	No	N/A	N/A	N/A	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.3.15291500-15292500	Fa H3K4me3-Fa H3K27Ac-Fb H4K20me1	TSS	Yes	Smp_123450	hypothetical protein	Peak size variations	Partial (not all marks)	Yes	Peak in Fa
Schisto_mansonl.Chr.3.15324000-15325500	Fa H3K4me3-Fa H3K27Ac-Fb H3K27me3	TSS	Yes	Smp_123440	cytoplasm protein	Peak size variations	No	Yes	Peak in Fa
Schisto_mansonl.Chr.3.15329000-15330000	Fa H3K4me3-Fb H4K20me1	TSS	Yes	Smp_123440	cytoplasm protein	Alternative TSS	Yes	Yes	Peak in Fa
Schisto_mansonl.Chr.3.17379000-17380000	Fa H3K4me3	Intergenic	No	N/A	N/A	N/A	Yes	Yes	Peak in Fa
Schisto_mansonl.Chr.3.21664500-21665500	Fa H3K4me3-Fa H3K27Ac	Promoter	Yes	Smp_023660	U3 small nucleolar ribonucleoprotein	Trans-spliced gene	Yes	Yes	Peak in Fa
Schisto_mansonl.Chr.3.21665500-21666500	Fa H3K27Ac-Fa H4K20me1	TSS	Yes	Smp_023660	U3 small nucleolar ribonucleoprotein	Trans-spliced gene, Alternative TSS	Yes	No	N/A
Schisto_mansonl.Chr.3.21665500-21668000	Fb H4K20me1-Fb H3K4me3	TSS	Yes	Smp_023660	U3 small nucleolar ribonucleoprotein	N/A	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.3.21669500-21670000	Fb H3K4me3-Fb H4K20me1	Intron 7	Yes	Smp_023660	U3 small nucleolar ribonucleoprotein	N/A	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.3.26146000-26146500	Fa H3K27Ac-Fb H3K27me3	Intron 1	Yes	Smp_197400	hypothetical protein	N/A	No	No	N/A
Schisto_mansonl.Chr.3.26155500-26156000	Fa H3K27Ac-Fb H3K27me3	Intergenic	No	N/A	N/A	N/A	No	No	N/A
Schisto_mansonl.Chr.3.26163000-26163500	Fa H3K27Ac-Fb H3K27me3	Intergenic	No	N/A	N/A	N/A	No	No	N/A
Schisto_mansonl.Chr.3.2648500-2649000	Fa H3K4me3-Fb H4K20me1	TSS	Yes	Smp_074060	Hypothetical protein	Trans-spliced gene, Alternative TSS	Yes	Yes	Peak in Fa
Schisto_mansonl.Chr.3.unplaced.SC_0083435000-14300	Fa H3K4me3-Fa H3K27Ac-Fb H3K27me3	TSS	Yes	N/A	N/A	N/A	Yes	Yes	Peak in Fa
Schisto_mansonl.Chr.4.12147000-12147500	Fa H4K20me1-Fb H3K27me3	TSS	Yes	Smp_042810	chromosome transmission fidelity protein 18	N/A	Partial (not all marks)	No	N/A
Schisto_mansonl.Chr.4.15299000-15299500	Fa H3K4me3-Fa H4K20me1-Fb H3K27Ac	Intron 2	Yes	Smp_059090	copropin 1 protein	N/A	No	Yes	Peak in Fa
Schisto_mansonl.Chr.5.681000-681500	Fa H3K4me3-Fa H3K27Ac-Fa H4K20me1-Fb H3K27me3	TSS	Yes	N/A	N/A	N/A	Partial (not all marks)	Yes	Peak in Fa
Schisto_mansonl.Chr.6.1534000-1535000	Fb H3K4me3-Fa H3K27me3	Intergenic	No	N/A	N/A	N/A	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.6.19516000-19516500	Fa H3K4me3-Fb H3K27me3	TSS	Yes	N/A	N/A	N/A	No	Yes	Peak in Fa
Schisto_mansonl.Chr.6.6013000-6013500	Fa H3K27me3-Fb H4K20me1	Intron 1	Yes	Smp_141770	glycerol kinase	Trans-spliced gene	No	No	N/A
Schisto_mansonl.Chr.W10.328000-30329000	Fa H3K4me3-Fa H3K27me3-Fb H3K27Ac	TSS	Yes	Smp_158770	calyphosin 2	Alternative TSS	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.W18553000-18554500	Fb H3K4me3-Fb H3K27Ac-Fb H4K20me1	TSS	Yes	N/A	N/A	Peak size variations	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.W31727000-37272000	Fb H3K4me3-Fb H4K20me1	TSS	Yes	N/A	N/A	N/A	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.W34360000-43604500	Fa H3K27me3-Fb H3K4me3	Intron 10	Yes	Smp_142680	long chain fatty acid coenzyme A ligase	N/A	No	Yes	Peak in Fb
Schisto_mansonl.Chr.W43606000-43607500	Fa H4K20me1-Fb H3K4me3	TSS	Yes	Smp_142680	long chain fatty acid coenzyme A ligase	Alternative TSS	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.W4719500-47222000	Fa H3K27Ac-Fa H4K20me1-Fb H3K27me3	TSS	Yes	N/A	N/A	N/A	No	No	N/A
Schisto_mansonl.Chr.W56510000-56511000	Fa H3K27me3-Fb H3K4me3-Fb H4K20me1	TSS	Yes	Smp_160320	DNA double strand break repair rad50 ATPase	Difference in H3K4me3	Yes	Yes	Peak in Fa
Schisto_mansonl.Chr.0080.245500-246000	Fa H4K20me1-Fb H3K27me3	TSS	Yes	N/A	N/A	N/A	No	No	N/A
Schisto_mansonl.Chr.032224500-225500	Fa H3K27me3-Fa H4K20me1	TSS	Yes	Smp_167110	Peptidase M8, leishmanolysin domain containing prote	Alternative TSS, Peak size variations	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.032226500-228500	Fa H3K27Ac-Fa H3K27me3-Fa H4K20me1	Intron 1	Yes	Smp_167110	Peptidase M8, leishmanolysin domain containing prote	Alternative TSS	Yes	No	N/A
Schisto_mansonl.Chr.044607000-6089500	Fa H3K27me3-Fa H4K20me1-Fb H3K4me3	TSS	Yes	Smp_162400	hypothetical protein	N/A	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.071253500-524000	Fa H3K27Ac-Fa H3K27me3-Fb H3K4me3	TSS	Yes	N/A	N/A	N/A	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.0713144000-144500	Fa H3K4me3-Fb H3K27me3	Intergenic	No	N/A	N/A	N/A	Partial (not all marks)	Yes	Peak in Fa
Schisto_mansonl.Chr.071321500-252500	Fb H3K27me3-Fb H3K27Ac	TSS	Yes	Smp_175670.1	egg protein CP391S	Peak size variations	Yes	Yes	Peak in Fb
Schisto_mansonl.Chr.027692500-932500	Fb H3K27Ac-Fb H4K20me1	Intron 1	Yes	Smp_175670.1	egg protein CP391S	Alternative splicing	Yes	No	N/A
Schisto_mansonl.Chr.027696000-97000	Fb H3K27Ac-Fb H4K20me1	TSS	Yes	Smp_175670.1	egg protein CP391S	Alternative splicing	Yes	No	N/A

Supplementary File 6: Repeated regions with histone mark enrichment between two samples

H3K4me3 Cercariae Bre Fa vs Cercariae Bre Fb

id	Mapped reads BreFa	Mapped reads BreFb	foldChange	Adjusted P-value
smr_8345_LTR (Transposable element)	168.575566522346	7.37746142090756	0.0437635273788603	5.99597209896279e-13
smr_1489_DNA (Transposable element)	152.709630849654	12.3318155001657	0.0807533580662415	6.52464456454301e-09
smr_2544_LINE/SmTRC1 (Transposable element)	529.525603076074	1227.44231330911	2.31800371158402	7.54353075877091e-05

H327Ac Cercariae Bre Fa vs Cercariae Gua Fa

id	Mapped reads BreFa	Mapped reads GuaFa	foldChange	Adjusted P-value
Sm_alphaT2 (Transposable element)	21092.1837344947	15693.8236536715	0.744058740015876	0.000600555710397419

H327Ac Adult Bre Fa vs Adult Gua Fa

id	Mapped reads BreFa	Mapped reads GuaFa	foldChange	Adjusted P-value
W12smr_315IHQ880216_SINE/W1 (Transposable element)	11868.9034552649	15481.2743816274	1.30435591122448	0.00086240684330716

H3K27me3 Adult Bre Fb vs Adult Gua Fb

id	Mapped reads BreFb	Mapped reads GuaFb	foldChange	Adjusted P-value
mr_2735_LTR (Transposable element)	945.095644481777	1304.25357231057	1.3800228367635	0.00225352808118762
smr_155_DNA/R2A_Nvi (Transposable element)	800.910151706166	1126.39558032615	1.40639443503945	0.00225352808118762
Sm_SR-A3 (Transposable element)	8567.07259420203	10708.7596085489	1.24999052953005	0.0191533762383679
Sm_SR2-A2 (Transposable element)	7989.72854654144	9985.80522989845	1.24983035052187	0.0191533762383679
R=130	3503.66522550793	4408.32669841616	1.25820431310673	0.0224367645896618
R=66	14168.8340832001	17407.3868596808	1.22856875572568	0.0362059871159524
smr_6170_DNA (Transposable element)	78.9491387639623	34.7062710805241	0.439602909213322	0.0362059871159524

H327me3 Cercariae Bre Fa vs Cercariae Gua Fa

id	Mapped reads BreFa	Mapped reads GuaFa	foldChange	Adjusted P-value
R=95	36205.2695200101	18644.175031674	0.51495749869697	0.00188836992602314
smr_80_LTR/SSU-rRNA_Dme (Transposable element)	1259.30495503892	606.000156631175	0.481217956148236	0.0086387844637136
smr_7429_DNA (Transposable element)	685.610383848251	308.455071542218	0.449898482883086	0.0134634060921523

H3K27me3 Cercariae Bre Fa vs Cercariae Bre Fb

id	Mapped reads BreFa	Mapped reads BreFb	foldChange	Adjusted P-value
smr_7429_DNA (Transposable element)	684.950007002065	419.477918809196	0.612421219827699	0.00209361203740669
smr_6364_LTR/TE-X-5_DR (Transposable element)	466.370372414642	724.307528685758	1.55307363316336	0.00651404624923826
W35smr_3102IHQ880241_LTR (Transposable element)	1542.14479517671	2031.09338519188	1.31705751077618	0.00997332880011734
R=882	845.107435109901	583.072548323701	0.689938964089076	0.0191360370104062

H4K20me1 Adult Bre Fb vs Adult Gua Fb

id	Mapped reads BreFb	Mapped reads GuaFb	foldChange	Adjusted P-value
smr_141_LTR/Helitron-2_SP (Transposable element)	229.524273357507	106.595973959637	0.46442135465822	5.91869595932002e-05
smr_4807_LTR (Transposable element)	344.843921899071	182.501057925718	0.529227996598217	0.000557501643399156

A) Differences between sympatric and allopatric condition with adjusted p-value < 0.01

id	Closest gene	Function (GeneDB)	Chromatin state*			
			baseMean	baseMeanA (sympatric)	baseMeanB (allopatric)	foldChange
XLOC_001340 Schisto_mansoni.Chr.0186:388,629-392,838			170.03982	300.6974335	39.38220663	0.1309695
XLOC_000588 Schisto_mansoni.Chr.143:39,772-43,223,933			323.48772	534.5747873	112.4005594	0.202616
XLOC_017374 Schisto_mansoni.Chr.0186:417,470-421,887	MEG-7		123.64456	210.0086993	36.82042837	0.1749948
XLOC_004124 Schisto_mansoni.Chr.1unplaced.SC_0153138:645-643,306			62.604335	12.1645361	13.04415363	0.162948
XLOC_018004 Schisto_mansoni.Chr.025:312-339	subfamily MMA unassigned peptidase		19.27942	36.99089822	1.567925622	0.0423868
XLOC_015630 Schisto_mansoni.Chr.0041:321,639-1,325,928	ncRNA		263.14178	402.9204983	123.3630634	0.306722
XLOC_001821 Schisto_mansoni.Chr.12:998,329-3,003,184			88.767308	140.0074467	37.52716836	0.2680369
XLOC_006198 Schisto_mansoni.Chr.3:25,682,852-25,687,344	Phospholipase C Like family member (p11)		70.099614	116.141977	24.0572508	0.2071366
XLOC_010417 Schisto_mansoni.Chr.6:5,373,130-5,377,513	radial spoke 3		402.82339	601.396156	204.2506142	0.3396274
XLOC_016058 Schisto_mansoni.Chr.013:671,014-675,167	hypothetical protein		223.89857	340.0744447	107.7272731	0.367622
XLOC_014981 Schisto_mansoni.Chr.W48:042,375-48,046,963	receptor type tyrosine protein phosphatase		73.770821	115.7446011	31.797000083	0.2747168
XLOC_0130570 Schisto_mansoni.Chr.167:054,192-54,754,309	unknown		108.75402	167.05431781	50.45313781	0.3020153
XLOC_006251 Schisto_mansoni.Chr.3:4,786,188-4,791,764			193.83314	289.1628636	98.50341854	0.3406053
XLOC_017621 Schisto_mansoni.Chr.021:267,370-272,822			41.989133	68.41271755	15.56554822	0.2275242
XLOC_017768 Schisto_mansoni.Chr.022:45,106-9,720			99.921151	149.951161	49.8918505	0.3327163
XLOC_01912 Schisto_mansoni.Chr.W10:42,277-10,147,280	hypothetical protein		87.165623	132.8478281	41.48341831	0.3022637
XLOC_010913 Schisto_mansoni.Chr.6:2,225,991-2,230,424			25.731382	42.95723067	8.505532507	0.198
XLOC_013002 Schisto_mansoni.Chr.3:701,062-370,310,666,931	hypothetical protein		35.995383	58.86690058	13.12386483	0.2222943
XLOC_008498 Schisto_mansoni.Chr.4:3,922,918-3,927,261			37.228341	59.26444254	15.19253911	0.256353
XLOC_002718 Schisto_mansoni.Chr.1:35,052,484-35,057,162	mitochondrial import inner membrane translocase		26.703385	66.42373366	18.12903618	0.2729301
XLOC_016348 Schisto_mansoni.Chr.0134:737,763-742,032			26.703335	46.53645252	6.870216904	0.1476309
XLOC_002288 Schisto_mansoni.Chr.122:291,720-22,296,892	Four and a half LIM domains protein 3		148.33342	213.930749	83.47377053	0.3915407

B) Loci that are at least 10fold overexpressed in the sympatric condition (allopatric repression)

id	Closest gene	Function	Chromatin state*			
			baseMean	baseMeanA (sympatric)	baseMeanB (allopatric)	foldChange
XLOC_018004 Schisto_mansoni.Chr.025:312-339			19.27942	36.99089822	1.567925622	0.0423868
XLOC_000588 Schisto_mansoni.Chr.021:617,116-161,548	subfamily MMA unassigned peptidase		7.0090764	0.892465864	0.0679938	0.0679938
XLOC_005077 Schisto_mansoni.Chr.2:362,466-366,463			9.5958003	17.89868686	1.272918014	0.07118
XLOC_009698 Schisto_mansoni.Chr.4unplaced.SC_0032:608,398-612,788			6.414694	11.9324729	0.892465864	0.074793
XLOC_015900 Schisto_mansoni.Chr.0080:439,444-443,972	acid spirngomyelinase phosphodiesterase 3b		8.5835646	15.9107531	1.256975774	0.0790046
XLOC_000588 Schisto_mansoni.Chr.14:327,862-43,242,098			5.815722	10.73907854	0.892465864	0.083045
XLOC_004140 Schisto_mansoni.Chr.1unplaced.SC_0153113:591,118,062			5.048175	16.70515649	1.404479578	0.0840746
XLOC_045033 Schisto_mansoni.Chr.W36:648,830-3,652,987	bone morphogenetic protein-like protein		5.418129	9.943792129	0.892465864	0.089751
XLOC_015715 Schisto_mansoni.Chr.W1unplaced.SC_0155:77432-581,561			0.080879	9.947079865	0.892465864	0.0897518
XLOC_017131 Schisto_mansoni.Chr.619:220,888-19,225,324	protocadherin 9		9.9748574	18.29634462	1.653370164	0.0903661
XLOC_003458 Schisto_mansoni.Chr.1163:070,504-63,074,789			7.1568881	13.1257043	1.187473472	0.0904693
XLOC_008987 Schisto_mansoni.Chr.4:6,295,146-6,299,377			6.0884121	11.13685453	1.039969668	0.0933809
XLOC_011310 Schisto_mansoni.Chr.618:515,802-18,520,117	serine/threonine protein kinase NIM1		5.2191984	9.545930973	0.892465864	0.0934918
XLOC_01725 Schisto_mansoni.Chr.7:9,490,319-9,494,669	hypothetical protein		9.3703709	17.10333956	1.637427924	0.0957375
XLOC_008874 Schisto_mansoni.Chr.41:077,899-1,082,127			6.7989023	12.33033135	1.187473472	0.0963373
XLOC_013410 Schisto_mansoni.Chr.W1:494,330-1,498,613	titin		8.8958826	10.39969668	1.0968387	0.0968387
XLOC_015851 Schisto_mansoni.Chr.W1unplaced.SC_0115:26,072-30,450			7.1993183	13.1257186	1.272918014	0.0969789
XLOC_005477 Schisto_mansoni.Chr.2:14,167,561-14,721,011	hook protein		8.2984669	15.1144572	1.48248708	0.0980837
XLOC_002936 Schisto_mansoni.Chr.143:839,298-43,843,576			6.5908273	10.34168503	1.039969668	0.0515609
XLOC_046229 Schisto_mansoni.Chr.W41:205,186-41,309,608			5.6908666	10.34403359	1.039969668	0.1005637

C) Loci that are at least 10fold overexpressed in the allopatric condition (allopatric induction)

id	Closest gene	Function (GeneDB)	Chromatin state*			
			baseMean	baseMeanA (sympatric)	baseMeanB (allopatric)	foldChange
XLOC_011340 Schisto_mansoni.Chr.619:502,578-19,507,707			15.651989	1.59021691	29.72495645	18.682936
XLOC_014883 Schisto_mansoni.Chr.W31:760,171-31,764,787	hypothetical protein		11.57688	1.93245703	21.96059407	18.404017
XLOC_017998 Schisto_mansoni.Chr.023:241,127-245,716	drug efflux protein		13.921689	1.591021691	26.25235692	16.500314
XLOC_009317 Schisto_mansoni.Chr.027:345,641-50,062	hypothetical protein		8.61587	1.93245703	16.03712833	13.439921
XLOC_008387 Schisto_mansoni.Chr.4:20,372,219-20,376,586	mitochondrial import inner membrane translocase		14.213273	1.988664897	26.4378817	13.294287
XLOC_013866 Schisto_mansoni.Chr.W15:857,070-15,861,676	poly(ADP-ribose) glycohydrolase		8.5126682	1.93245703	15.83209074	13.268089
XLOC_017249 Schisto_mansoni.Chr.0186:107,836-177,192	TBC1 domain family 3		10.16124	1.5909553	18.75152432	11.73759
XLOC_017303 Schisto_mansoni.Chr.0183:70,163-174,738	poly(ADP-ribose) polymerase		7.500208	1.93245703	13.8077025	11.571104
XLOC_002524 Schisto_mansoni.Chr.128:815,499-28,820,755	zinc finger protein 542		12.444342	1.988781808	22.89990022	11.54537
XLOC_017630 Schisto_mansoni.Chr.W7:231,556-723,877			14.701763	2.386491405	27.0703395	11.320818
XLOC_002578 Schisto_mansoni.Chr.417:655,397-659,688	TNF receptor associated protein 1		7.3425958	1.93245703	13.41914584	11.0693
XLOC_018481 Schisto_mansoni.Chr.0394:115,34-16,071			9.7093743	1.5909553	17.8277934	11.20576
XLOC_015644 Schisto_mansoni.Chr.77:62,256-766,611	E3 ubiquitin protein ligase NEDD4		7.1183729	1.93245703	13.0450009	10.9311
XLOC_015408 Schisto_mansoni.Chr.0013:2,600,821-3,285,069	splicing factor 3a subunit 2		7.092758	1.93245703	12.99270589	10.888542
XLOC_008877 Schisto_mansoni.Chr.41:133,253-1,137,133	transmembrane protein tmem1b		6.938557	1.93245703	10.629721	10.629721
XLOC_005452 Schisto_mansoni.Chr.0037:1,233,027-2,128,462	prospero homeobox protein 2		34.621632	5.966240385	63.270791958	10.605838
XLOC_007393 Schisto_mansoni.Chr.3:176,970-476,179,518	ATP dependent Clp Protease ATP binding subunit		9.9812346	1.591005189	16.37146346	10.290009
XLOC_005190 Schisto_mansoni.Chr.2:4,708,105-4,712,578	permease 1 heavy chain		6.7252724	1.93245703	12.25729919	10.27234
XLOC_002909 Schisto_mansoni.Chr.143:177,168-43,882,969	hypothetical protein		6.7064635	1.93245703	12.21968137	10.40708
XLOC_009051 Schisto_mansoni.Chr.4:9,312,794-9,317,094	RUN domain containing protein 1		11.145851	1.988781808	20.30291926	10.208721
XLOC_016129 Schisto_mansoni.Chr.0125:198,039-202,1560			6.5965776	1.93245703	11.99990094	10.056528

* Ep/Cseq chromatin states

** chromatin states were only established for contigs assembled on chromosome level

IV.3. RÉSULTATS ET DISCUSSION SUPPLÉMENTAIRES NON PUBLIÉS

IV.3.1. Effet de la souche de *B. glabrata* sur le succès de greffe

Lors de l'expérience de greffe de sporocystes, une quinzaine de mollusques *BgBre* et *BgGua* ont été greffés pour chacun des deux génotypes du parasite (Fa et Fb). Deux mois plus tard, les mollusques ont été criblés pour estimer le succès de greffe. Il a été remarqué dans notre échantillon que la technique de greffe présentait de meilleurs résultats chez les mollusques allopatriques *BgGua* (**Tableau 1**). Cette observation s'est montrée statistiquement significative lors d'un test exact de Fisher (p-value = 0,036).

Tableau 1 : Efficacité des greffes de sporocystes et de l'infection des hôtes vertébrés.

Génotype <i>S. mansoni</i>	<i>B. glabrata</i> Bre		<i>B. glabrata</i> Gua		Adultes obtenus des souris	
	Infecté	Non infecté	Infecté	Non infecté	De <i>BgBre</i>	de <i>BgGua</i>
Femelle A (Fa)	1	13	6	10	17	30
Femelle B (Fa)	5	10	9	8	171	405
Total	6	23	15	18	188	435

La technique de greffe a été employée dans les années 80 pour maintenir des clones mâles ou femelles sur une longue période. À l'époque, il n'était pas encore possible de faire la détermination sexuelle de clones cercaires par PCR sur des marqueurs génétiques comme c'est le cas maintenant (Beltran *et al.* 2008). Comme seuls les vers adultes présentent un dimorphisme sexuel, il fallait infecter des souris avec les cercaires issues d'une infection mono-miracidiale et les perfuser 30 jours plus tard pour déterminer le sexe (Duvall & DeWitt 1967). La greffe de sporocystes permettait alors de conserver les clones mâles et femelles en les greffant de mollusques en mollusques de quinze à vingt fois (Anne Rognon, communication personnelle). Les greffes d'une souche de mollusque vers une autre n'avaient jamais été expérimentées. Lors d'une infection classique avec *SmBre*, la prévalence sur les mollusques sympatriques *BgBre* est de 50-100% et de 30-85% en allopatrique sur *Bgua*, en fonction du nombre de miracidia utilisés (Theron *et al.* 2014). En nous basant sur ces chiffres, nous présumons que le succès serait moindre en greffant un sporocyste développé sur un mollusque sympatrique vers un mollusque allopatrique. Comme on peut l'observer dans le **Tableau 1**, l'inverse s'est produit. Un seul échantillon de mollusques greffés incite toutefois à rester prudent sur la généralité de cette observation. Nous n'avons pas effectué d'expériences supplémentaires pour mieux comprendre ce phénomène, mais nous avons soulevé une hypothèse. En conditions naturelles, la

compatibilité entre *S. mansoni* et *B. glabrata* implique de nombreuses interactions moléculaires entre des produits de sécrétions libérés immédiatement après la pénétration du miracidium (ex : *Schistosoma mansoni* Polymorphic Mucins, SmPoMucs), et des molécules de reconnaissance présentes dans l'hôte (ex : *Fibrinogen RElated Proteins*, FREPS) (Guillou et al. 2007; Roger et al. 2008). Ces produits de sécrétions et ces interactions jouent un rôle crucial dans le succès ou l'échec de l'infection. Les sporocystes greffés ne libèrent pas ces produits de sécrétions tels que le feraient des miracidia (Guillou et al. 2007) et la reconnaissance du pathogène se déroule en impliquant probablement d'autres mécanismes, qui, à notre connaissance, n'ont pas été étudiés. Cependant, des travaux du laboratoire ont montré que les mollusques *BgGua* produisent naturellement moins d'espèces réactives de l'oxygène (ROS) (Guillou et al. 2007) et moins de lectines (Richard Gallinier, communication personnelle), deux types de molécules impliquées dans la destruction du parasite, que les *BgBre*. Il est donc envisageable que cette différence moléculaire explique le meilleur succès de greffe chez les *BgGua*. Il pourrait aussi s'agir simplement d'une différence physiologique, par exemple une différence souche spécifique au niveau de la cicatrisation des tissus du mollusque après la greffe qui aurait un impact sur l'implantation du sporocyste.

Le passage par l'hôte intermédiaire semble avoir un effet sur l'intensité de l'infection chez l'hôte vertébré. Pour les deux génotypes du parasite (Fa & Fb), et les deux conditions environnementales (allopatrique et sympatrique), trois souris ont été infectées par 300 cercaires. Bien qu'il n'y ait pas de significativité statistique, on remarque une tendance lors des infections des hôtes vertébrés (*M. musculus*) à obtenir plus de vers adultes à partir des cercaires issues des hôtes allopatriques. On peut émettre l'hypothèse que les sporocystes et cercaires issues des mollusques *BgGua*, ayant des mécanismes de défenses moins importants que les *BgBre* (ROS et lectines), investissent moins de ressources dans leur survie dans l'hôte intermédiaires et plus dans des mécanismes leur offrant un meilleur succès d'infection chez l'hôte vertébré. On peut également proposer l'hypothèse que la pénétration dans l'hôte intermédiaire soit le filtre le plus important pour l'infection de l'hôte intermédiaire, et qu'une fois la pénétration réussie (naturellement ou artificiellement par la greffe), le développement suit son cours normal.

IV.3.2. Effet du génotype du parasite sur l'intensité de l'infection chez la souris

Toujours dans le **Tableau 1**, on observe une différence marquante de l'intensité de l'infection (nombre de vers adultes récoltés) chez l'hôte vertébré entre les génotypes Fa et Fb. Douze fois plus de schistosomes adultes ont été récupérés pour le génotype Fb. Encore une fois, avec un seul échantillon, il est risqué de généraliser, mais cette tendance est intéressante. Comme nous disposons des génomes et d'une partie des épigénomes (pour quatre modifications d'histones), cela ouvre la perspective d'étudier plus en détail quels pourraient être les déterminants moléculaires impliqués dans le succès d'infection de l'hôte vertébré. Une meilleure caractérisation des gènes associés aux différences épigénétiques entre les deux génotypes est en cours. Les différences génétiques ont été identifiées et sont ponctuelles à l'exception des modifications du nombre de copies (CNV) observées sur des locus spécifiques (*cf.* article soumis dans **Molecular Ecology**), ce qui offre déjà des régions candidates à étudier.

IV.4. PERSPECTIVES

Ces travaux ont montré que certaines marques épigénétiques (celles étant aléatoires ou reliées au génotype) peuvent être héritées somatiquement entre les stades cercaires et adultes. Dans la continuité de cette étude d'héritabilité, j'ai refait une expérience de greffe de sporocystes, encore une fois en regardant le développement de *SmBre* dans des mollusques *BgBre* et *BgGua*. Cette fois-ci, j'ai utilisé un génotype mâle et un génotype femelle, de manière à pouvoir obtenir une descendance sexuée. Dans cette expérience, j'ai récolté cercaires, adultes (mâles et femelles séparément) et miracidia. Ces échantillons biologiques pourront être utilisés par le laboratoire pour mieux caractériser de possibles différences épigénétiques entre femelles et mâles (aux stades cercaires et adultes), et observer dans quelle mesure des épimutations chez cercaires et adultes peuvent être transmises aux miracidia.

Si elle n'est pas écologiquement réaliste, la greffe de sporocystes offre une bonne opportunité d'observer comment l'épigénome de *S. mansoni* évolue lorsqu'il se développe dans hôtes de souches distinctes. Cette technique me semble prometteuse pour étudier plus en détail les interactions et la compatibilité entre le parasite et son hôte intermédiaire.

Lors des expériences de compatibilité, pour mesurer la prévalence ou l'intensité par exemple (comme dans Theron *et al.* (2014)), la compatibilité n'est jamais totale en infections mono-miracidiales. En fonction des souches de parasites et de mollusques utilisés, une certaine proportion de mollusques reste non-infectée. S'il est éventuellement possible d'arriver à identifier les déterminants permettant la compatibilité chez les deux partenaires, lorsque l'infection échoue, on ne peut pas regarder du côté du parasite incompatible (vu que les miracidia ne sont pas clonaux et qu'on ne peut pas à la fois comparer leur génome/epigénome et les utiliser dans une infection). En suivant la logique des hypothèses classiques de compatibilité *S. mansoni* / *B. glabrata*, (Roger *et al.* 2008; Moné *et al.* 2010; Mitta *et al.* 2012), on peut supposer qu'un même miracidia a un potentiel d'infection qui varie en fonction de sa souche et de la souche de l'hôte intermédiaire. Un miracidia ne peut pas être utilisé pour infecter deux mollusques différents, mais je pense envisageable d'utiliser la greffe de sporocyste comme un proxy à cet effet.

Dans le travail présenté dans ce chapitre, l'étape limitante des analyses a été la comparaison des profils épigénétiques entre conditions environnementales, génotypes et individus. À l'heure actuelle, les logiciels permettant ce type d'analyses sont peu nombreux et encore assez expérimentaux. Dans ce contexte, j'effectue une collaboration avec Aaron Taudt et le groupe de Maria-Colomé Tatché (*Laboratory of Quantitative Epigenetics*, Groningen, Pays-Bas) pour tester un logiciel de leur conception sur mes données. Ce logiciel est encore en développement, mais donne des résultats prometteurs. Nous sommes actuellement en train de procéder à une caractérisation de la structure chromatinienne chez cercaires et adultes, avec une résolution beaucoup plus fine que ce qui a été présenté dans le chapitre II de ce manuscrit. Nous espérons publier un article sur ces travaux au cours de l'année 2016.

CHAPITRE V

TECHNIQUES D'ÉTUDE DE L'ÉPIGÉNOME DE POCILLOPORA DAMICORNIS

V.1. MISE EN CONTEXTE

Le projet d'étude de la thermotolérance sur le corail tropical *P. damicornis* a été initié (et est toujours coordonné) par Jérémie Vidal-Dupiol, en 2008. Dans cette expérience, des boutures de coraux issus d'une même colonie mère (donc en théorie, génétiquement identiques) ont été soumis à un événement annuel d'élévation de température écologiquement réaliste, durant trois ans (Vidal-Dupiol 2011). D'une année à l'autre, les coraux se sont montrés plus résistants au stress thermique, blanchissant de plus en plus tardivement. Au fil de l'expérience (avant, pendant et après chaque stress thermique), des échantillons biologiques ont été collectés pour étudier différents aspects de la réponse du corail aux changements de température : transcriptome, génome, épigénome et composition des clades de *Symbiodinium* dans les tissus (**Figure 28**). Ce projet est toujours en cours, et a permis le séquençage et l'annotation du génome de *P. damicornis* par notre équipe, qui devrait bientôt être publié.

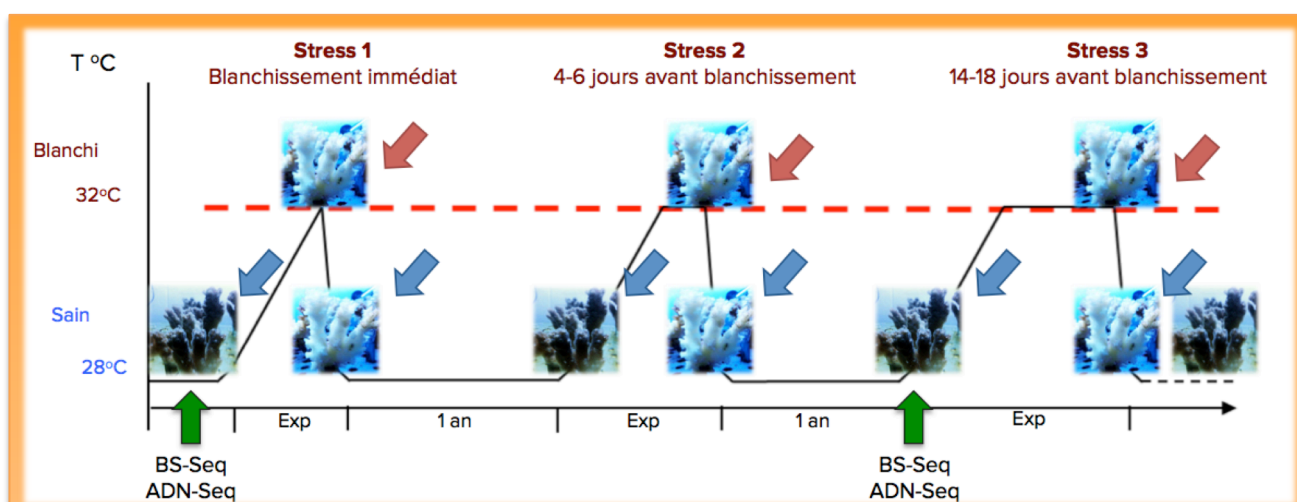


Figure 28 : Schéma expérimental d'induction de thermotolérance chez *P. damicornis*. Des colonies clonales saines ont été soumises à un stress thermique écologiquement réaliste induisant le blanchissement trois années successives. Le temps de résistance au blanchissement s'accroît après chaque stress. Les **flèches vertes** représentent les points pour lesquels l'analyse du génome (DNA-Seq) et du méthylome (BS-Seq) a été faite. Transcriptome (RNA-Seq) et composition en clade de *Symbiodinium* ont été testés aux points marqués par des **flèches bleues**, et transcriptome seul pour les **flèches rouges**. Photographies de Kelly Brener-Raffalli.

Mon implication au sein de ce projet s'est portée sur l'optimisation des techniques épigénétiques sur *P. damicornis*. Jusqu'à aujourd'hui, il n'existe aucune publication présentant des résultats expérimentaux sur la structure chromatinienne des coraux hermatypiques, que ce soit au niveau de la méthylation de l'ADN ou des modifications d'histones. Il était donc nécessaire de mettre au point les techniques moléculaires

permettant l'étude de l'épigénome de *P. damicornis*. Une première étape a consisté à développer un protocole permettant d'isoler les noyaux des cellules de corail des *Symbiodinium* (qui sont intracellulaires). La séparation a été un succès. D'autres expériences ont ensuite été effectuées pour mieux caractériser la structure chromatinienne du corail, et il a été décidé de se concentrer sur la méthylation des cytosines plutôt que sur les modifications d'histones, pour des considérations techniques. L'ADN de trois réplicats (trois boutures, mais issues de la même colonie mère) avant le premier et troisième stress (**Figure 28, flèches vertes**) a été converti au bisulfite puis séquencé sur technologie Illumina. Les séquences ont été obtenues au printemps 2015 et j'ai encadré une stagiaire de Master I, Ariadna Picart Picolo, qui s'est très fortement impliquée dans l'établissement d'un premier profil du méthylome global de *P. damicornis*.

V.2. PUBLICATION ASSOCIÉE

Une synthèse décrivant la méthode permettant l'isolation des noyaux cellulaires de *P. damicornis*, et une première caractérisation de la structure chromatinienne et du méthylome a été faite sous forme d'article. Sa vocation est d'être un article compagnon lors de la publication du génome de *P. damicornis*, au courant de l'année 2016. Pour cette raison, il n'a pas été soumis à une revue scientifique pour le moment, mais est présenté ici sous une forme quasi-définitive.

Tropical coral *Pocillopora damicornis* has a methylome similar to other invertebrates, but displays an unusual chromatin structure

David Roquis^{*1,2}, Ariadna Picart Picolo^{*1,2}, Céline Cosseau², Christoph Grunau² & Jeremie Vidal-Dupiol^{1,2}

^{*}Both authors contributed equally

¹ Université de Perpignan Via Domitia, Perpignan, F-66860, France,

² CNRS, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE), Perpignan, F-66860, France,

Abstract

Pocillopora damicornis is a hermatypic coral with a world-wide distribution a strong ecological importance. It is threatened by anthropic perturbations, and principally global warming. Thermal stress can induce bleaching, a phenomenon in which coral polyps expel their endosymbiotic unicellular algae, and can lead to the death of entire colonies. Previous works have shown that genetically identical colonies can have different level of thermotolerance depending on the environmental conditions they previously faced. Epigenetic modifications are good candidates to explain this phenomenon. The clonal nature of a colony and the possibility of generating new genetically identical colonies through propagation make corals an attractive model to study the impact of the environment on the epigenome. However, until now, no work has been published on *P. damicornis* (and corals in general) epigenome. One of the main problems is caused by the intracellular location of *Symbiodinium*, which makes complicated to isolate coral chromatin and DNA without contamination from endosymbiotic biological material. Here, we propose a simple method to purify *P. damicornis* chromatin. We also provide the first description of a coral methylome, with methylation principally in CpG context, and with a genome wide mosaic pattern. We also observe an unusual, uncharacterized chromatin structure never described before in other *Cnidaria* or invertebrates.

1. Introduction

Coral reefs are important and complex ecosystems with a worldwide distribution, principally in tropical and are subtropical latitudes. They are biodiversity hotspots, housing over 900,000 species (Hughes *et al.* 2002), and are central to the economy of many countries for their role in fishery and tourism (Hughes *et al.* 2002). Central to these ecosystems are colonial, calcifying organisms known as hermatypic corals (*Cnidaria* phylum). Colonies are composed of genetically identical individual, the polyps. They live in symbiosis with dinoflagellate microalgae from the *Symbiodinium* genus, which provide through photosynthesis most the energy needed for the coral to grow and build their aragonite skeleton (Cooper *et al.* 2011). These corals are threatened by anthropic perturbations, and principally global warming, which disturbs their relationship with the

Symbiodinium. Through a phenomenon commonly named bleaching, coral colonies respond to stress by expelling their endosymbiotic dinoflagellate. Depending on their species, polyps can survive a few days to few weeks through heterotrophy. Without a quick return to normal environmental conditions allowing the recolonizations by the *Symbiodinium*, colonies will ultimately perish, leading to the desertification of the reef (Douglas 2003). Bleaching events have been more and more frequent in the past decades, mostly driven by the increase in water temperature, and their impacts on the reef ecosystems are a major source of concern for ecologist (Douglas 2003; Buddemeier *et al.* 2004; Hoegh-Guldberg 2011). However, recent studies have shown that, at least in some coral species, acclimatization or fast adaptation occurs and leads to a heat resistance phenotype. Thermotolerant corals resist longer and to higher water temperature without bleaching (Maynard *et al.* 2008; Palumbi *et al.* 2014), but are also susceptible to lose their resistance if they are moved to environment with optimal conditions and low temperature fluctuations for a couple of years (Palumbi *et al.* 2014). The molecular mechanisms underlying thermotolerance are to be discovered, and a better knowledge of them is essential to develop efficient protection policies for coral reefs.

Epigenetic modifications are good candidates to explain rapid, inheritable and reversible phenotypes without change in the DNA Sequence. They range from chemical modifications of DNA (eg. cytosine methylation), covalent changes of proteins interacting with DNA (e.g. histone modifications), as well as nuclear localization of chromosomes, and short untranslated RNA involved in post-transcriptional silencing of genes and repeated regions (Wu & Morris 2001). These modifications have an impact on the chromatin, leading to transcriptionally permissive or repressive structures (euchromatin and heterochromatin, respectively). It is now clear that environmental factor can influence the epigenome and induce phenotypes (Feil & Fraga 2012; Duncan *et al.* 2014). In natural populations with genetic diversity, disentangling the respective roles of genetics and epigenetics in environmentally triggered phenotypes is not feasible. Corals are therefore excellent models in that aspect due to the clonal nature of the colonies. It is easy to generate new colonies through propagation (cutting a coral branch and fixing it somewhere else) and observe how their epigenome changes under various environmental conditions, while maintaining the genotype constant. Despite their ecological and experimental attractiveness, technical challenges have until now prevented corals to serve as model species to study the effect of environment on epigenome. Literature describing the epigenome of hermatypic corals is scarce, and we found only two publications focused on this matter (Liew *et al.* 2014; Dixon *et al.* 2014). Tissues and cells extraction from the stony skeleton and working at physiological, marine, salt concentration to maintain cell integrity and chromatin structure possible, but the main problem is the separation of coral and *Symbiodinium* biological material. *Symbiodinium* are endocellular, and to focus on the coral epigenetic modifications, it is necessary to use an extraction technique that would lysate coral cells, but not *Symbiodinium*, and then separate coral nuclei from cell debris and endosymbiontes.

In this article, we propose a method we developed to isolate coral nuclei of *P. damicornis*, a specific hermatypic coral, and provide a first insight about the chromatin structure of this species. We chose this coral for its worldwide distribution and ecological importance (IUCN), its fast growth rate allowing to quickly generate new colonies through propagations (Hughes *et al.* 2015), and documented thermotolerance (Vidal-Dupiol *et al.* 2009; 2014). *P. damicornis* has a genome size of approximately 325 mb (Vidal-Dupol *et al.* under preparation), while *Symbiodinium* genome size is estimated at 3,000 mb (Shoguchi *et al.* 2013), which means that a single *Symbiodinium* brings as much DNA as ten coral cells, hence the need to reduce contamination from endosymbiotic material at the lowest possible level. We used these isolated coral nuclei for the analysis of epigenetic information carriers. We show that mosaic cytosine methylation occurs and that 5-methyl-cytosine is predominantly found in CpGs. We also found that canonical histone modifications exist, but that *P. damicornis* has an unusual chromatin structure.

2. Material & Methods

2.1. Biological material

Healthy *Pocillopora damicornis* isolates used in the present study were harvested in Lombok (Indonesia, CITES number: 06832/VI/SATS/LN/2001) and maintained at the Banyuls Aquarium (France) under optimal conditions. Bleached coral colonies were obtained through an ecologically realistic thermal stress, as described in (Vidal-Dupiol *et al.* 2009). Fragment of bleached colonies were collected 6 to 10 days after the beginning of bleaching, when there is an estimated reduction of 99% of *Symbiodinium* in coral tissues, but before any necrosis could be observed on the colonies. Healthy and bleached coral fragments were immediately frozen and stored in liquid nitrogen.

In this work, we also employed *Aiptasia pallida* and *Daphnia magna* for comparison in the chromatin extraction and digestion experiments. *A. pallida* is from the same *Cnidaria* phylum as *P. damicornis*, shares an endosymbiotic relationship with *Symbiodinium*, but does not have an aragonite skeleton. They were acquired from the Banyuls Aquarium (France). *Daphnia magna* is a freshwater microinvertebrate phylogenetically unrelated to *A. pallida* and *P. damicornis*, but on which we previously optimized the chromatin digestion technique. Prof. Ellen Decaestecker from KU Leuven (Belgium) graciously provided them to us.

Aposymbiotic *Aiptasia* were obtained through a menthol treatment, used to artificially induce bleaching in *Cnidaria* (Wang *et al.* 2012). Anemones were maintained in 4 L of artificial seawater (37 g/L of Reef Crystals) with the addition of 18 mL of a solution of 20 g of menthol /L of absolute ethanol, in the absence of light for four days. Final menthol concentration in the seawater is of 0.58 mmol /L. Water was replaced everyday, always supplemented with the same concentration of menthol. Healthy and bleached *Aiptasia* were conditioned separately in Eppendorf tubes (0.180g per tube) and immediately frozen and stored in liquid nitrogen.

Symbiodinium cultures were obtained from the Roscoff Culture Collection (strain CCMP 2246, cat# RCC4017).

2.2. In-Silico identification of canonical histones

We were interested to use isolated coral chromatin to perform chromatin immunoprecipitation (ChIP) on histone modifications. Before starting experiments, we wanted to confirm that histones were present in *P. damicornis* genome, and that their amino acid sequence would be similar enough to other metazoan to be able to target them with commercial antibodies. We searched the *P. damicornis* transcriptome (Vidal-Dupiol *et al.* 2013; 2014) for transcripts annotated as histones H2A, H2B, H3, H4. These sequences were translated to amino-acid sequences and aligned with MEGA (Tamura *et al.* 2013) to reference protein sequences from *Mus musculus*, *Schistosoma mansoni*, a platyhelminth for which we have extensive experience in ChIP, and two other *Cnidaria*: *Hydra vulgaris* and *Nematostella vectensis*. Accession numbers for each species is in **Table 1**.

Table 1: List of accession numbers used to compare *P. damicornis* canonical histone sequences.

	H2A	H2B	H3	H4
<i>N. vectensis</i>	XM_001623180	XM_001620416	XM_001639326	XM_006544099
<i>H. vulgaris</i>	XM_002158289	XM_002156049	XM_002164185	XM_002161956
<i>M. musculus</i>	M33988	NM_175663	AC034285	XP_001630070
<i>S. mansoni</i>	HE601626	HE601624	CCD77737	AAG25601

2.3. Western blots

As we had the intention to perform chromatin immunoprecipitation targeting histone modifications, we first performed western blots using commercial anti-histones antibody to assess their efficiency and specificity on *P. damicornis* canonical and modified histones. Bleached and healthy coral fragment of approximately 5 mm of diameter (~ 0.5 g) were put in 1.5 mL tubes filled with 500 µL of a solution containing 62.5 mM TRIS/Cl pH 6.8, 3% SDS, 10% sucrose, 0.2M dithiotreitol (DTT) and 1.25 mM sodium butyrate. Tubes were sonicated using a Vibra Cell 75185 at 70% amplitude, 3 times 15 seconds, on ice. A cooldown on ice of 30 seconds was done between each sonication. Dinoflagellata phylum, which includes *Symbiodinium*, does not possess histones (Rizzo 2003) and can be used as a negative control. We chose hamster (*Mesocricetus auratus*) as a positive control, as most antibodies are commercially developed and tested on mammals. Hamster brain and cultured aposymbiotic *Symbiodinium* protein extracts were processed the same way as *P. damicornis*.

The extract was cleared by centrifugation for 30 minutes at 1500 g, and the supernatant was collected. Total proteins (5 mg per sample) were mixed with Laemmli buffer (final concentration 1X) and denatured at 99°C for 5 mins. ECL Full-Range Rainbow (Amersham cat# RPN800E) was used as a molecular weight marker. Protein separation was done on 10% SDS-PAGE gel electrophoresis before being blotted on a nitrocellulose

membrane (Trans-Blot turbo, Bio-Rad). The membrane was blocked with 5% non-fat dry milk in TBST (TBS buffer containing 0.05% tween 20) one hour at room temperature. One of the following primary antibody, diluted in 5% non fat milk in TBST was used for each western blot: anti-histone H3 (Abcam cat# ab1791 lot# 784471, dilution 1/500), anti-H3K36me2 (Abcam cat# ab1220 lot# GR75522, dilution 1/500), and anti-H3K27me3 (Diagenode, cat#C15410069, lot# A1821D). After incubation, membrane was washed 3 times for 10 minutes in TBST. It was incubated with secondary antibody (peroxidase conjugated, goat purified anti-rabbit IgG [Pierce cat# 31460, lot# HB987318]) diluted 1/5000 in TBST for 1 hour. After washing 3 times for 10 minutes in TBST, the detection was carried out using the ECL reagents and the ChemiDoc MP Imaging system (BioRad).

2.4. Coral nuclei isolation

Our chromatin extraction protocol is based on (Cosseau & Grunau 2011), but had to be adapted so that salinity, ionic strength and osmolarity of various buffers would match those of seawater. *P. damicornis* polyps (healthy or bleached) were removed from the aragonite skeleton using an Airpick. This instrument is a modified version of Waterpick (Vidal-Dupiol *et al.* 2011) designed to gently blow air rather than liquid. Advantage of the Airpick compared to the Waterpick is that it can be used with smaller volume of extraction solution and with a lower amount of biological material. Extractions with Airpick were performed on ice for 5 minutes (or until no tissues were left on the skeleton) with the coral fragments (approximately 5 mm of diameter and 0.5 g each) inside a 50 mL tubes filled with 10 mL of a buffer 1 (See **Table 2** for composition of all the buffers used in this article). Tubes were then centrifuged at 800 g at 4°C for 10 minutes. Supernatant was carefully removed, and pellets were gently resuspended in 1 mL of buffer 1 and 1 mL of buffer 2. The mix was transferred in Dounce homogenizer and ground on ice for 4 minutes with pestle A, and then left to rest on ice for 7 minutes. Liquid was transferred on corex tubes already containing 8 mL of buffer 3, in a manner that the homogenate would form a layer on top of buffer 3. The two solutions have a different density, which allow them to stay one on top of the other without mixing. Corex tubes were centrifuged 20 mins at 4°C and 7,800 g, with the lowest break possible. This step allows separating *P. damicornis* nuclei, which will form a pellet at the bottom of the tube, from cell debris and *Symbiodinium*, which stay at the interphase in the homogenate on top of buffer 3. Supernatant was completely removed by pouring out of the tubes and then by aspiration with micropipette. Pellets, containing coral nuclei and chromatin, were either used DNA purification (to use for bisulfite sequencing) or chromatin digestion (with the objective of using digested chromatin for ChIP).

200 µg of healthy and bleached *Aiptasia* samples were directly ground in Dounce homogeneizer with 1 mL of buffers 1 & 3, and then processed the same way as *P. damicornis*. For *Daphnia*, 70 frozen adults were processed exactly as described in (Cosseau & Grunau 2011).

Table 2: Final concentration for all components of the various buffers used in the chromatin extraction and micrococcal nuclease digestion.

	Buffer 1	Buffer 2	Buffer 3	Mnase Buffer
KCl	30 mM	30 mM	30 mM	
NaCl	500 mM	500 mM	500 mM	500 mM
MgCl ₂	2.5 mM	2.5 mM	2.5 mM	40 mM
Tris/Cl pH 7.4	15 mM	15 mM	15 mM	10 mM
Sucrose	1 M	1 M	1.5 M	1 M
Sodium butyrate	5 mM	5 mM	5 mM	5 mM
Dithiotreitol (DTT)	25 mM	25 mM	25 mM	
Phenylmethanesulfonyl fluoride	0.5 mM	0.5 mM	0.5 mM	0.1 mM
cOmplete protease inhibitor tablet (Roche cat #11 836 145 001)	1 X	1 X	1 X	
NP-40		0,3 %		
CaCl ₂				10 mM

2.5. Chromatin digestion

One of the most important steps to study histone modifications through chromatin immunoprecipitation is chromatin shearing. This can be done either with crosslinking followed by sonication, or natively using micrococcal nuclease (MNase). Crosslinking creates covalent bounds between DNA and proteins interacting with it, and can be broken in di or tri nucleosomal fragments through timed sonication. This method is usually employed for transcription factor analysis, but is not needed in the case of histones, as they are strongly bound to DNA. Crosslinking can also cause a shift in nucleosomes position. We favored the native approach with MNase, a bacterial enzyme digesting DNA between nucleosomes. Digestion time has to be optimized to obtain only mono- to penta-nucleosomal fragments. To do so, nuclei pellets (bleached and healthy coral, *A. pallida* and *D. magna*) from step 2.4 were resuspended in 1.5 mL of MNase digestion buffer (see **Table 2** for composition). Twenty microliters of the solution were kept for a coral nuclei and *Symbiodinium* count (See step 2.5). Aliquots of 250 µL were prepared in Eppendorf tubes for digestion. 1 µL of 1 U/µL MNase (Affymetrix cat# 70196Y) was added to each tube, and incubation at 37°C was done for 0, 2, 4, 6 and 8 minutes. We also prepared a negative control with no MNase and incubated at 37°C for 8 minutes. Digestion was stopped with the addition of 20 µL of 1M EDTA and tubes were left on ice for 5 minutes. Chromatin was purified using QIAquick PCR Purification kit (Qiagen cat# 28104) and DNA fragments were eluted in 30 µL of TE buffer (supplied with the kit). All of the purified DNA fragments were then separated on a 1.8 % agarose gel stained with ethidium bromide for 30 minutes at 100 V (in 1X Tris Borate EDTA buffer).

2.5. Estimation of *P. damicornis* nuclei and intact *Symbiodinium* in chromatin pellets

The chromatin extraction method we used is gentle enough to break *P. damicornis* cytoplasmic membrane without lysing their nuclei or *Symbiodinium*. Coral nuclei are then isolated from cell debris and *Symbiodinium* due to the density difference of the solutions used during the centrifugation step in corex tubes (step 2.4). To get an estimation of the amount of possible *Symbiodinium* contamination in the nuclei/chromatin pellet, we took 20 µL of coral nuclei resuspended in MNase digestion buffer (from step 2.5) and added Hoechst 33342 (Invitrogen cat# C10329) at a final concentration of 1/2000. We put 10 µL on a microscope slide and observed on a fluorescent microscope at 350 nm. Hoechst 33342 stains coral nuclei with a blue fluorescence, but does not enter intact *Symbiodiniums* (red fluorescence due to chlorophyll).

2.6. DNA purification and bisulfite sequencing

Bisulfite sequencing needs DNA to be completely protein-free. To do so, nuclei pellets obtained at step 2.4 were resuspended in 180 µl of the ATL buffer from the DNeasy Blood and Tissues Kit (Qiagen cat# 69504) and then processed as described in the manufacturer documentation. We sent the purified DNA from three healthy coral colonies for bisulfite sequencing (BS-Seq) to GATC biotech (www.gatc-biotech.com), which performed both bisulfite conversion and sequencing. Libraries with fragment size ranging from 150 to 500 bp were paired-end sequenced (100 bp on each extremity) on a Illumina Hi-Seq 2000.

2.7. Quality control, alignment on *P. damicornis* genome and methylome description

All data treatment was carried out under a local galaxy instance (Goecks *et al.* 2010) (<http://bioinfo.univ-perp.fr>). Fastq Groomer v1.0 was used for verification of the fastqsanger format, and the FASTX-Toolkit v0.0.13 (Compute quality statistics, Draw quality score boxplot, Draw nucleotides distribution chart) was used for initial quality control. First and last five nucleotides for all reads were removed using the TRIM tool from the FASTX-Toolkit because of lower quality. Overall, read quality was judged sufficiently good (the majority of reads showed a quality score above 26 for the non-trimmed position) and no further quality filter was applied.

Read alignment on the *P. damicornis* reference genome (Vidal-Dupiol *et al.* in preparation) was done in paired-end using Bismark v2.74 (Krueger & Andrews 2011) and Bowtie2 (Langmead & Salzberg 2012) with default parameters. We only kept alignments for which both read mates were successfully aligned and fragment size was between 150 and 500 bp. BAM files output from Bismark were sorted and converted to SAM format using Samtools v0.1.19 (Li *et al.* 2009). SAM files were loaded into the R package methylKit v0.5.6 (Akalin *et al.* 2012) using the `read.bismark()` command and specifying the cytosine context (CpG, CHG or CHH, respectively). Afterwards, we used the `getMethylationStats()` command to obtain the percentage of methylated cytosines per context. Visualization of the

262 methylated regions was done by converting alignment BED files into bigWig format with the
 263 Wig/Bedgraph-to-bigWig converter v1.1.1 included in Galaxy (Goecks *et al.* 2010). BigWig
 264 where then uploaded into Trackster, the visual analysis environment embed in Galaxy
 265 (Goecks *et al.* 2010).

267 3. Results

268 3.1. Coral nuclei are isolated with less than 2% contamination of *Symbiodinium*

269 Observation under fluorescent microscope showed that coral nuclei were not degraded
 270 and that there was little cell debris. We counted the number of intact *Symbiodinium* (which
 271 have a red fluorescence at 350 nm because of their chlorophyll content) and intact coral
 272 nuclei (with blue fluorescence caused by Hoechst 33342 binding to DNA). We observed the
 273 result of 3 different nuclei isolation and never found more than 2 intact *Symbiodinium* for 98
 274 coral nuclei. The same result was obtained with healthy *A. pallida*. As expected, in bleached
 275 coral or *Aiptasia* extracts, no *Symbiodinium* were found. Alignment of the BS-Seq reads also
 276 confirm the low amount of contamination from *Symbiodinium*, as less than 5% of the reads
 277 (per sample) mapped on *Symbiodinium* genome. An average of 56.4 % of the reads
 278 obtained for our three BS-Seq samples (healthy colonies) mapped on *P. damicornis*
 279 reference genome. This may look like a small percentage of alignment, but it is what is
 280 classically observed with bisulfite sequencing (Gavery & Roberts 2010).

282 3.2. Mosaic DNA methylation occurs principally in CpG context in *P. damicornis*

283 Within our three replicates from healthy coral colonies, analysis of the bisulfite
 284 sequencing data through methylKit informed us that 0.98% of cytosines are methylated
 285 genome-wide. Methylation occurs principally in CpG dinucleotide context, with 3.70 % of
 286 them being methylated. It is also found in the two other possible nucleotide contexts, with
 287 0.20 % of the CHG and 0.20 % of the CHH being methylated. Methylation is distributed in a
 288 mosaic fashion, with large methylated intragenic regions (mostly on exons) interspersed
 289 with large unmethylated regions.

291 3.3. Histones are present in *P. damicornis*, but chromatin displays an unusual structure 292 not found in another *Cnidaria*.

293 Transcripts for the four histones composing nucleosomes were found in *P. damicornis*
 294 transcriptome and were aligned to other species to see similarity in amino acid sequences.
 295 H3 and H4, have 94% and 98% of identity in their amino acid sequences between *P.*
 296 *damicornis* and *M. musculus*. Other two histones inside nucleosomes, H2a and H2b, have
 297 respectively 92% and 87% of identity with *M. musculus*. Histone tails for H3 and H4, where
 298 chemical modification occurs, were always 100% identical between mice and *P. damicornis*.

Sequence identity for the four histones was similar with *S. mansoni*, but increased when comparing with less phylogenetically distant organisms (*H. vulgaris* and *N. vectensis*).

We then performed western blots to see if commercial antibodies targeting histone modifications could be used on *P. damicornis*. We tested hamster brain (positive control), aposymbiotic *Symbiodinium* (negative control, as dinoflagellata have no histones (Rizzo 2003)), bleached and healthy corals. Results of the western blots can be seen on **Figure 1**. Results for the three antibodies (anti-H3, anti-H3K27me3, anti-H3K36me2) were consistent between organisms. In all three cases, no signal was detected for *Symbiodinium* and a signal at the expected size (approximately 15 kDa) for histone H3 was observed in hamster. In healthy coral, a signal was observed, but at 12 kDa. Bleached coral displayed both the 12 kDa band and the expected 15kDa band.

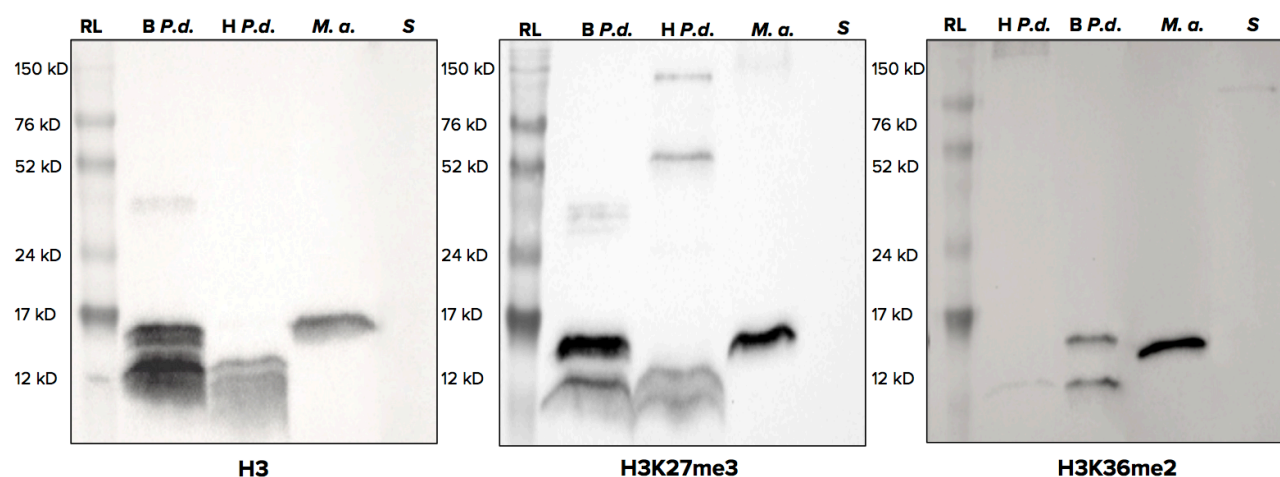


Figure 1: Western blots using antibodies targeting histone H3 (left), trimethylation of lysine 27 of histone H3 (H3K27me3, center) and déméthylation of lysine 36 of histone H3 (H3K36me2, right). RL = Amersham ECL Rainbow Ladder; B *P.d.* = Bleached *P. damicornis*; H *P.d.* = Healthy *P. damicornis*; S = *Symbiodinium*; M. a. = *M. auratus*.

From these results, we suspected that there could be a different chromatin structure between bleached and healthy *P. damicornis*, and we carried on with chromatin MNase digestion. MNase digests DNA between nucleosomes, which creates DNA fragments with sizes that are multiples of approximately 150 bp (length of DNA wrapped around a nucleosome). When separated on a gel, it looks like a ladder. After prolonged digestion, only 150 bp fragments, corresponding to a mono-nucleosomal digestion, would be detectable on an agarose gel. We did not observe this type of result for neither healthy nor bleached coral. Instead DNA is completely degraded with time, and no nucleosomal size fragments are seen. To verify that it was not a problem with the chromatin extraction and digestion technique, we performed the same experiment on healthy and bleached *Aiptasia pallida*, a cnidarian also sharing an endosymbiotic interaction with *Symbiodinium*, and *Daphnia magna*, a freshwater crustacean. As it can be seen on **Figure 2**, digestion profile in the shape of a “ladder” is observed for both of them, but not in *P. damicornis*.

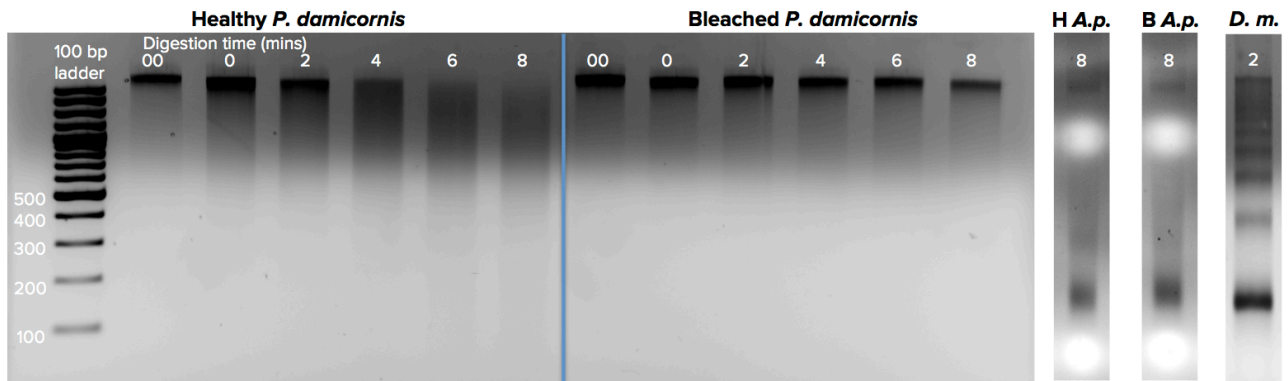


Figure 2: MNase digestion profile of chromatin from healthy and bleached *P. damicornis* in function of digestion time at 37°C. 00 corresponds to chromatin left at 37°C for 8 minutes, without the addition of MNase, and 0 is addition of MNase immediately followed by incubation on ice with EDTA (enzyme inhibitor). We also show almost complete digestion with fragments around 150 bp in healthy (H A.p.) and bleached (B A.p.) *A. pallida*. Partial digestion, with the typical ladder shape, can be seen in *D. magna* (*D. m.*) on the far right.

4. Discussion

In this article, we propose a simple method to isolate *P. damicornis* chromatin and DNA from *Symbiodinium*'s, and provide the first description of a coral epigenome. We observed a “mosaic” methylation pattern, with methylation-free regions interspersed with intragenic highly methylated ones (mostly on exons). Around 1% of the total cytosines are methylated, principally in the dinucleotide CpG. CHG and CHH methylation occurs nonetheless, but a lower frequency. These characteristics are in line with other invertebrates in which DNA methylation is present and a methylome was described (Suzuki & Bird 2008; Gavary & Roberts 2010; Xiang *et al.* 2010; Lyko *et al.* 2010; Flores & Amdam 2011; Fneich *et al.* 2013). Further analysis of the relation of DNA methylation to other elements of the genome (genes, exons, transcription start and end sites etc.) will be possible after complete annotation of the genome.

We intended to use the isolated coral nuclei to perform chromatin immunoprecipitation (ChIP) on histone modifications. Alignments of protein sequences of histones H2A, H2B, H3 and H4, which compose nucleosomes, show that they are well conserved in *P. damicornis*, with only little difference with mice. However, we do not observe the histone H3 band at the expected size (15 kDa) in healthy coral for any of the three antibodies we used (anti-H3, anti-H3K27me3, anti-H3K36me2). Instead, a single band of lower molecular weight (12 kDa) is visible. Bleached coral are different, with both the expected H3 band at 15 kDa, and the truncated one at 12 kDa. As anticipated, no signal was detected in *Symbiodinium* extract, as they do not possess histones (Rizzo 2003). *Symbiodinium* live within coral cell, and their presence or absence may cause major cellular and nuclear reorganization. There is the possibility that coral chromatin may be compacted in different fashion, with other histone forms or histone-like proteins, depending on its symbiotic state. *Symbiodinium* also live within *A. pallida* cells, but we did not observe a different chromatin structure between symbiotic and aposymbiotic anemone. Further analyses are needed to confirm this hypothesis.

We were unable to digest the chromatin into nucleosomal fragments. MNase cuts preferentially between nucleosomes because of steric hindrance. DNA is wrapped around nucleosomes for a 150 bp length, and incomplete digestion present a “ladder” profile on gels, with each bands being a multiple of 150 bp, while complete digestion shows a unique, intense band at 150 bp (Hewish & Burgoyne 1973; Noll & Kornberg 1977; Wu *et al.* 1979; Hörz & Altenburger 1981; Keene & Elgin 1981; Zaret 2005). In *P. damicornis*, either healthy or bleached, DNA is completely digested by MNase (**Figure 2**), similarly to what would happen in absence of nucleosomes (Hewish & Burgoyne 1973). This was different in *A. pallida*, which, like *P. damicornis*, is also a symbiotic *Cnidaria*. Since the only difference between the procedures used for *P. damicornis* and *A. pallida* is the removal of the aragonite skeleton with the Airpick, we also tried other types of extraction with Waterpik, cryogenic grinding or manual polyp removal. Neither of the protocols delivered expected chromatin profiles (data not shown). In conclusion, our data suggest that *P. damicornis* has a different chromatin structure from other metazoa.

5. References

- Akalin A, Kormaksson M, Li S *et al.* (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, **13**, R87.
- Buddemeier RW, Kleypas JA, Aronson RB, (2004) *Coral reefs and global climate change* Pew Center on Global Climate Change.
- Cooper TF, Lai M, Ulstrup KE *et al.* (2011) *Symbiodinium* genotypic and environmental controls on lipids in reef building corals. *PLoS ONE*, **6**, e20434.
- Cosseau C, Grunau C (2011) Native chromatin immunoprecipitation. *Methods in molecular biology (Clifton, N.J.)*, **791**, 195–212.
- Dixon GB, Bay LK, Matz MV (2014) Bimodal signatures of germline methylation are linked with gene expression plasticity in the coral *Acropora millepora*. **15**, 1–11.
- Douglas AE (2003) Coral bleaching—how and why? *Marine Pollution Bulletin*, **46**, 385–392.
- Duncan EJ, Gluckman PD, Dearden PK (2014) Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **322**, 208–220.
- Feil R, Fraga MF (2012) Epigenetics and the environment: emerging patterns and implications. *Nature Reviews Genetics*, **13**, 97–109.
- Flores KB, Amdam GV (2011) Deciphering a methylome: what can we read into patterns of DNA methylation? *Journal of Experimental Biology*, **214**, 3155–3163.
- Fneich S, Dheilly N, Adema C *et al.* (2013) 5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of *Biomphalaria glabrata*, a snail intermediate host of *Schistosoma mansoni*. *Parasites & Vectors*, **6**, 1–1.
- Gavery MR, Roberts SB (2010) DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). *BMC Genomics*, **11**, 483.

- 406 Goecks J, Nekrutenko A, Taylor J, Galaxy Team T (2010) Galaxy: a comprehensive approach
407 for supporting accessible, reproducible, and transparent computational research in the life
408 sciences. *Genome Biology*, **11**, R86.
- 409 Hewish DR, Burgoyne LA (1973) Chromatin sub-structure. The digestion of chromatin DNA
410 at regularly spaced sites by a nuclear deoxyribonuclease. *Biochemical and biophysical
411 research communications*, **52**, 504–510.
- 412 Hoegh-Guldberg O (2011) Coral reef ecosystems and anthropogenic climate change.
413 *Regional Environmental Change*, **11**, 215–227.
- 414 Hörz W, Altenburger W (1981) Sequence specific cleavage of DNA by micrococcal nuclease.
415 *Nucleic Acids Research*, **9**, 2643-2658.
- 416 Hughes RN, Hughes DJ, Smith IP, Dale AC (2015) *Oceanography and marine biology: an
417 annual review*. CRC Press.
- 418 Hughes TP, Bellwood DR, Connolly SR (2002) Biodiversity hotspots, centres of endemism,
419 and the conservation of coral reefs. *Ecology Letters*, **5**, 775–784.
- 420 IUCN *Pocillopora damicornis* (2014): Hoeksema, B.W., Rogers, A. & Quibilan, M.C. Available
421 at <http://www.iucnredlist.org/details/133222/0>.
- 422 Keene MA, Elgin SC (1981) Micrococcal nuclease as a probe of DNA sequence organization
423 and chromatin structure. *Cell*, **27**, 57–64.
- 424 Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-
425 Seq applications. *Bioinformatics*, **27**, 1571–1572.
- 426 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature
427 Methods*, **9**, 357–359.
- 428 Li H, Handsaker B, Wysoker A et al. (2009) The Sequence Alignment/Map format and
429 SAMtools. *Bioinformatics*, **25**, 2078–2079.
- 430 Liew YJ, Aranda M, Carr A et al. (2014) Identification of MicroRNAs in the Coral Stylophora
431 pistillata (M Medina, Ed.). *PLoS ONE*, **9**, e91101.
- 432 Lyko F, Foret S, Kucharski R et al. (2010) The honey bee epigenomes: differential
433 methylation of brain DNA in queens and workers (L Keller, Ed.). *PLoS Biology*, **8**, e1000506.
- 434 Maynard JA, Anthony KRN, Marshall PA, Masiri I (2008) Major bleaching events can lead to
435 increased thermal tolerance in corals. *Marine Biology*, **155**, 173–182.
- 436 Noll M, Kornberg RD (1977) Action of micrococcal nuclease on chromatin and the location of
437 histone H1. *Journal of molecular biology*, **109**, 393–404.
- 438 Palumbi SR, Barshis DJ, Traylor-Knowles N, Bay RA (2014) Mechanisms of reef coral
439 resistance to future climate change. *Science (New York, N.Y.)*, **344**, 895–898.
- 440 Rizzo PJ (2003) Those amazing dinoflagellate chromosomes. *Cell Research*, **13**, 215–217.
- 441 Shoguchi E, Shinzato C, Kawashima T et al. (2013) Draft assembly of the *Symbiodinium
442 minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology*, **23**, 1399–
443 1408.
- 444 Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from
445 epigenomics. *Nature Reviews Genetics*, **9**, 465–476.

- 446 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary
447 Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.
- 448 Vidal-Dupiol J, Adjerdoud M, Roger E *et al.* (2009) Coral bleaching under thermal stress:
449 putative involvement of host/symbiont recognition mechanisms. *BMC Physiology*, **9**, 14.
- 450 Vidal-Dupiol J, Dheilly NM, Rondon R *et al.* (2014) Thermal stress triggers broad *Pocillopora*
451 *damicornis* transcriptomic remodeling, while *vibrio coralliilyticus* infection induces a more
452 targeted immuno-suppression response (P Jiravanichpaisal, Ed.). *PLoS ONE*, **9**, e107672.
- 453 Vidal-Dupiol J, Ladriere O, Meistertzheim AL *et al.* (2011) Physiological responses of the
454 scleractinian coral *Pocillopora damicornis* to bacterial stress from *Vibrio coralliilyticus*.
455 *Journal of Experimental Biology*, **214**, 1533–1545.
- 456 Vidal-Dupiol J, Zoccola D, Tambutte E *et al.* (2013) Genes related to ion-transport and
457 energy production are upregulated in response to CO₂-driven pH decrease in corals: new
458 insights from transcriptome analysis (S Dupont, Ed.). *PLoS ONE*, **8**, e58652.
- 459 Wang J-T, Chen Y-Y, Tew KS, Meng P-J, Chen CA (2012) Physiological and biochemical
460 performances of menthol-induced aposymbiotic corals (CR Voolstra, Ed.). *PLoS ONE*, **7**,
461 e46406.
- 462 Wu CT, Morris JR (2001) Genes, genetics, and epigenetics: a correspondence. *Science*
463 *(New York, N.Y.)*, **293**, 1103–1105.
- 464 Wu C, Bingham PM, Livak KJ, Holmgren R, Elgin S (1979) The chromatin structure of specific
465 genes: I. Evidence for higher order domains of defined DNA sequence. *Cell*, **16**, 797-808.
- 466 Xiang H, Zhu J, Chen Q *et al.* (2010) Single base-resolution methylome of the silkworm
467 reveals a sparse epigenomic map. *Nature Biotechnology*, **28**, 516–520.
- 468 Zaret K (2005) Micrococcal nuclease analysis of chromatin structure. In: *Current protocols in*
469 *molecular biology*, 21.1.1-21.1.17. John Wiley & Sons, inc.

V.3. RÉSULTATS ET DISCUSSION SUPPLÉMENTAIRES NON PUBLIÉS

V.3.1. Optimisation de la technique d'immunoprécipitation de chromatine

Beaucoup de temps et d'énergie ont été consacrés à l'optimisation de la technique d'immunoprécipitation de chromatine (ChIP) sur *P. damicornis*. De part l'absence d'histones chez *Symbiodinium* (Rizzo 2003), cela semblait une voie intéressante pour étudier l'épigénome de *P. damicornis* si la méthode d'isolation des noyaux s'était avérée inefficace. Le pipeline d'optimisation du ChIP sur histone proposé par Cosseau & Grunau (2011) est très performant pour mettre au point cette technique sur d'autres organismes. Dans le cadre de ma thèse, il m'a permis de rapidement développer le ChIP sur le crustacé *Daphnia magna* (sous forme de collaboration), l'anémone de mer *Aiptasia pallida* (cf. article présenté dans ce chapitre) et le nématode *M. incognita* (voir publication en **Annexe 2**). Toutefois, cela n'a pas été possible chez *P. damicornis*, malgré de nombreuses tentatives.

Une cinquantaine d'essais (**Annexe 3**) ont été effectués, sur des variations du protocole de Cosseau & Grunau (2011), mais également en testant d'autres protocoles, dont un utilisé avec succès sur *Nematostella vectensis*, un autre cnidaire marin (Schwaiger *et al.* 2014). Des approches natives ou avec *crosslink* (c'est-à-dire formation de liens covalents entre l'ADN et les histones), ont été expérimentées, mais sans amélioration des résultats. Le même profil de digestion que celui présenté dans l'article est systématiquement obtenu lorsque la fragmentation de la chromatine se fait par digestion à la nucléase micrococcale (MNase). Ce profil correspond à celui obtenu lorsque de l'ADN sans histone est digéré par MNase (**Figure 29**) (Hewish & Burgoyne 1973).

Que ce soit par *crosslink* suivi de sonication, ou de manière native, quelques unes de nos tentatives ont été menées jusqu'au bout du protocole de ChIP, mais nous n'avons jamais réussi à obtenir de l'ADN immunoprécipité. Ces résultats renforcent l'idée qu'il existe une structure chromatinienne atypique chez *P. damicornis*, mais pas universelle à tous les cnidaires. Un test additionnel, en natif, a été tenté sur une autre espèce de corail assymbiotique, *Coralium rubrum*. Un profil de digestion similaire à *P. damicornis* a été obtenu. Toutefois, la très faible quantité de matériel biologique de *C. rubrum* utilisée, et l'absence de réplicats, m'invite à une grande prudence quant à l'interprétation de ce résultat. Il serait pertinent de tester plus d'espèces de coraux, hermatypiques ou non, et de

voir dans quelle mesure cette structure chromatinienne inhabituelle est répandue. L'absence d'outils moléculaires chez *P. damicornis* rend difficile l'étude de sa structure chromatinienne *in vivo*, mais il serait envisageable d'effectuer des approches d'immunomarquages des histones sur des tissus de polypes fixés pour mieux comprendre leur dynamique dans le noyau. Une approche protéomique (par spectrométrie de masse par exemple) permettrait peut-être d'élucider pourquoi on observe des profils de western blots différents pour l'histone 3 entre coraux sains et blanchis.

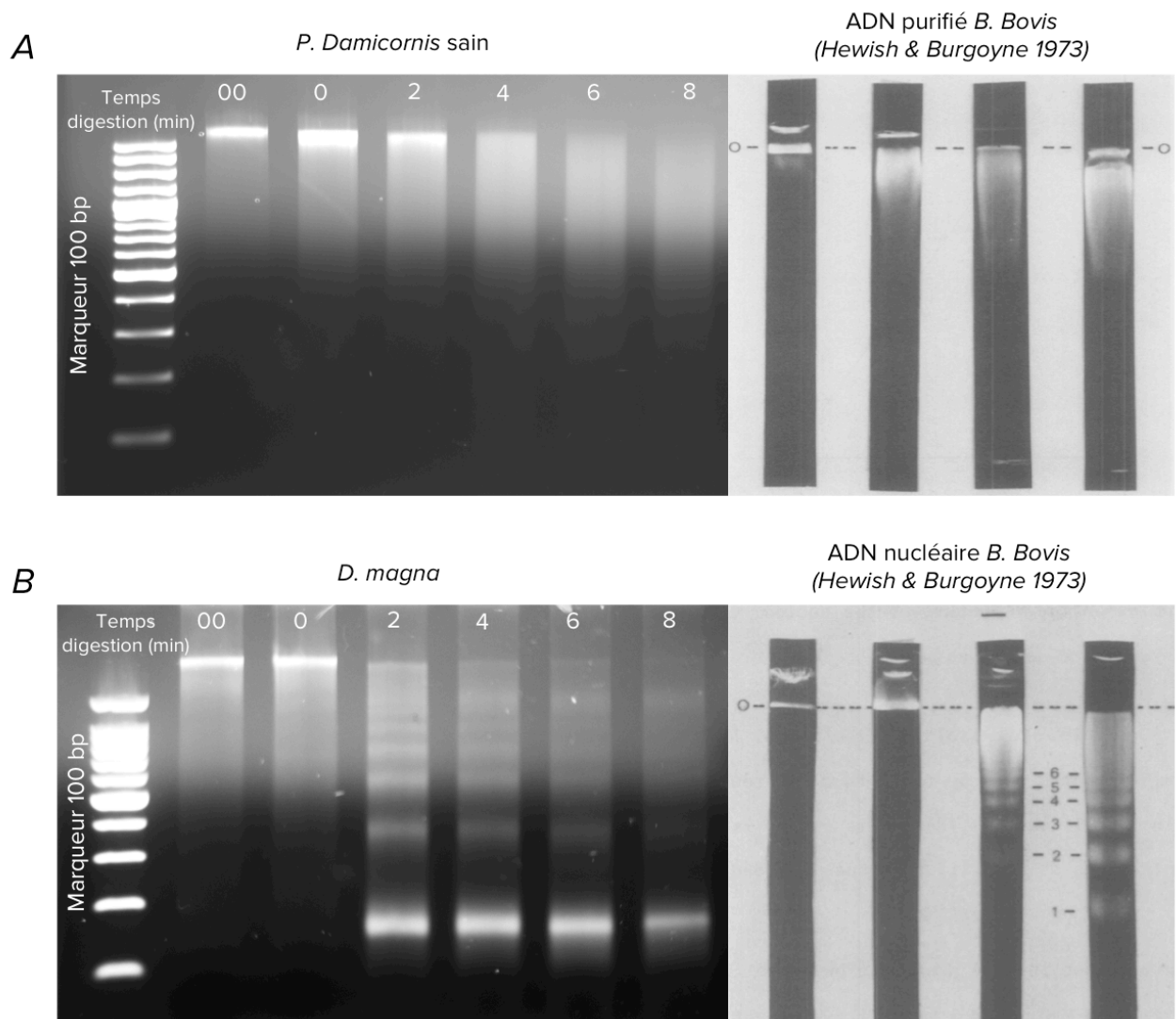


Figure 29 : (A) Comparaison d'un exemple de profil de digestion à la MNase obtenu sur *P. damicornis* avec celui publié par Hewish & Burgoyne (1973) sur de l'ADN purifié. (B) Comparaison d'un profil de digestion à la MNase obtenu avec le crustacé d'eau douce *Daphnia magna* avec celui publié par Hewish & Burgoyne (1973) sur de l'ADN nucléaire contenant encore ses histones. Les digestions de *D. magna* et *P. damicornis* sont effectuées à 37°C et stoppées par ajout d'EDTA et mises sur glace. Le temps 00 correspond à un échantillon contrôle mis à 37°C pendant 8 minutes, mais sans ajout de MNase. Le temps 0 correspond à un échantillon contrôle pour lequel la MNase et son inhibiteur (EDTA) ont été ajoutés en même temps, et immédiatement mis sur glace.

V.3.2. Comparaison des méthylomes entre coraux thermotolérants et sensibles

L'analyse des données de méthylation entre coraux avant le premier stress et avant le dernier stress se poursuit actuellement. Des résultats préliminaires de l'analyse comparative ont déjà identifié 2 898 régions du génome différentiellement méthylées (un exemple est visible sur la **Figure 30**). Ces régions ont une longueur d'au moins 300 paires de bases et ont au minimum 20% de cytosines méthylées en plus ou en moins entre les deux conditions. Les détails sur la détection de ces régions sont disponibles dans le rapport de stage de d'Ariadna Picart-Piccolo, étudiante de Master I que j'ai encadrée (Picart Piccolo 2015). La prochaine étape consistera en une caractérisation de ces régions pour déterminer ce qu'elles contiennent (gènes? Éléments répétés? Éléments régulateurs?) et observer s'il existe une corrélation entre ces régions et les différences transcriptomiques.

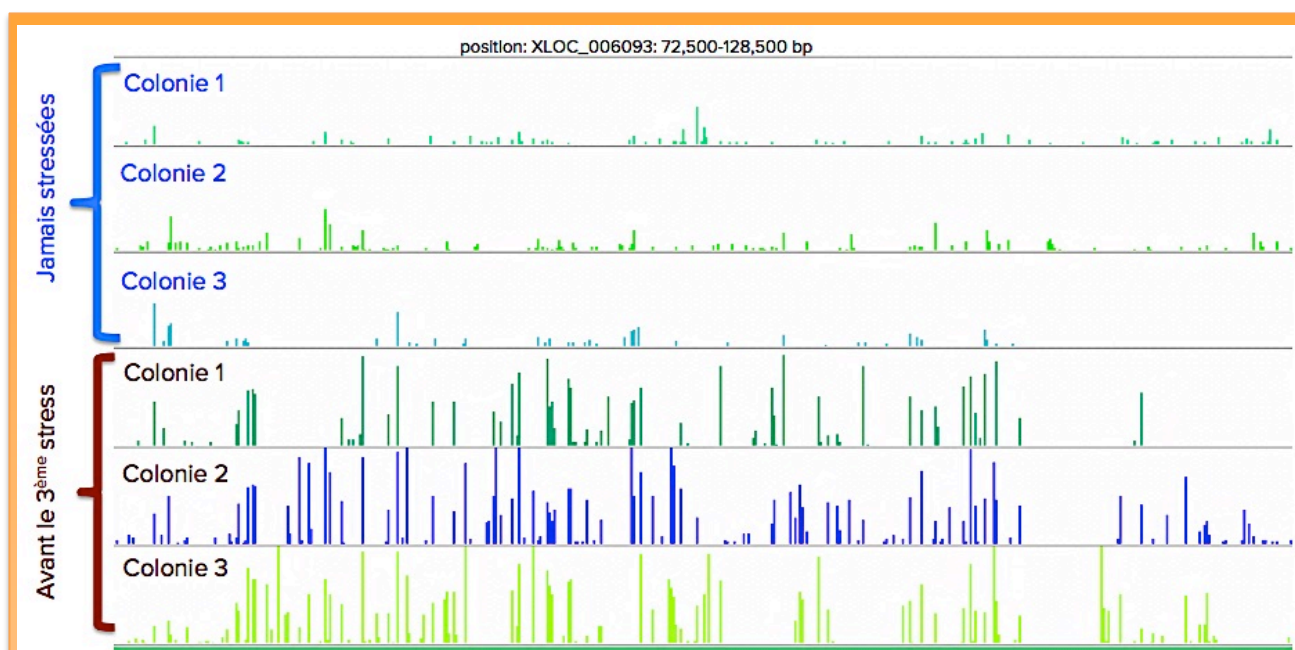


Figure 30 : Représentation graphique d'une région différentiellement méthylée entre coraux jamais stressés et coraux stressés deux fois. Les barres de couleurs symbolisent le ratio de cytosine méthylées / non-méthylées à une position génomique donnée. Plus la barre est haute, plus le taux de m5C est élevé. On observe que le pourcentage de méthylation est plus élevé dans cette région chez les colonies ayant été stressées plusieurs fois.

CHAPITRE VI

DISCUSSION

VI.1. RAPPEL DES PRINCIPAUX RÉSULTATS

Au début de ma thèse, je me suis fixé comme objectif scientifique principal d'explorer quels pouvaient être les liens existant entre *Dauermodifikationen* et mécanismes épigénétiques. J'ai divisé cet objectif en deux parties, que l'on peut résumer en deux grands questionnements : **(I)** l'environnement induit-il des changements de la structure chromatinienne ciblés, aléatoires, ou les deux, et **(II)** dans quelle mesure ces changements sont-ils héréditaires mitotiquement et/ou méiotiquement? J'ai mené mes investigations sur deux modèles aux extrêmes des relations symbiotiques : le corail mutualiste *P. damicornis* et le parasite *S. mansoni*. Ces deux organismes présentent de bonnes caractéristiques biologiques pour être en mesure de répondre au mieux à mes deux questionnements : présence de *Dauermodifikationen*, phase de reproduction clonale permettant de mieux faire les liens entre phénotype/épigénétique/environnement, modèles en interactions symbiotiques connus pour répondre rapidement aux changements environnementaux, peu ou pas de variabilité héréditaire de type sociale ou culturelle. Ces deux espèces, contrairement à beaucoup d'organismes modèles en biologie, ont une importance sanitaire et médicale (*S. mansoni*) ou environnementale et économique (*P. damicornis*). Ces deux espèces ont toutefois l'inconvénient d'être moins bien connues sur le plan épigénétique que d'autres organismes modèles (*D. melanogaster*, *M. musculus*, *A. thaliana*, etc.). La première partie de mon travail a consisté à mieux caractériser leurs épigénomes respectifs.

VI.1.1. Description des épigénomes de *S. mansoni* et *P. damicornis*

VI.1.1.1. Structure chromatinienne différente entre *S. mansoni* cercaires et adultes

Chez *S. mansoni*, nous avons détaillé la distribution génomique de six modifications d'histones chez cercaires et adultes, et deux sur schistosomules. Nous avons montré que certaines marques ont des profils similaires aux autres métazoaires connus (H3K4me3, H3K9ac, H3K27me3 et H4K20me1) et d'autres présentent des spécificités uniques au parasite (H3K9me3, H3K27ac). Nous avons aussi remarqué une structure chromatinienne particulière, avec la présence de quatre marques (H3K4me3, H3K9ac, H3K9me3 et H3K27me3) aux sites d'initiation de la transcription (*transcription start site*, TSS) de gènes chez les cercaires. Par des expériences de marquage de la synthèse d'ARN nous avons constaté que la présence de ces quatre marques au TSS est associée à une suspension de

l'activité transcriptomique chez les cercaires. Nous avons aussi mis en évidence des changements majeurs de structure chromatinienne lors du passage du stade cercaire à adulte. Ces bouleversements s'initient très rapidement après pénétration de la cercaire dans l'hôte vertébré et se caractérisent par la reprise quasi immédiate de la transcription (observée dès 2h après transformation en schistosomule).

Résultats clefs :

- ✓ *H3K9me3 et H3K27ac ont des profils uniques à S. mansoni.*
- ✓ *Structures chromatiniennes très différentes entre cercaires et adultes.*
- ✓ *Structure chromatinienne stoppant la transcription chez les cercaires.*

VI.1.1.2. Premier méthylome d'un corail scléractiniaire

L'épigénome de *P. damicornis* est encore à un stade exploratoire, mais nos travaux ont permis de développer les outils moléculaires menant à une première description de son méthylome. Le pourcentage de cytosines méthylées sur l'ensemble du génome a été mesuré à 0,98%. Cette méthylation survient principalement dans un contexte de dinucléotides CpG, 3,70% d'entre eux sont méthylés sur l'ensemble du génome. On trouve également une faible proportion de méthylation dans les deux autres contextes possibles, avec 0,20% des cytosines en CHG et 0,20% en CHH. La distribution de la 5-méthylcytosine (5mC) est sous forme mosaïque, avec des régions intragéniques (principalement au niveau des exons) méthylées, entrecoupées de larges régions non-méthylées. Ces caractéristiques du méthylome sont similaires à ce qui a déjà été observé chez d'autres invertébrés chez lesquels on retrouve la 5-mC (Suzuki & Bird 2008; Gavery & Roberts 2010; Xiang *et al.* 2010; Lyko *et al.* 2010; Flores & Amdam 2011; Fneich *et al.* 2013). Dans notre approche d'optimisation de la technique d'immunoprécipitation de chromatine (ChIP), nous avons constaté une structure chromatinienne inhabituelle. Bien que des transcrits et des western blots ont montré la présence d'histones chez *P. damicornis*, nous n'avons pas été en mesure de digérer la chromatine en fragments mononucléosomiaux malgré de nombreuses tentatives. Cela a pourtant été possible très facilement chez *Aiptasia pallida*, une autre espèce de cnidaire symbiotique. Cela signifie que ce type de structure chromatinienne singulière n'est pas répandu à toutes les espèces de ce phylum. Il serait intéressant de regarder chez d'autres cnidaires pour déterminer si cette caractéristique est unique chez *P. damicornis* ou distribuée chez d'autres coraux. Les western blots suggèrent qu'il existe

deux formes d'histones H3, distribuées différemment entre coraux blanchis et sains. Les *Symbiodinium* étant localisés à l'intérieur des cellules du corail, il est raisonnable de penser que leur présence/absence influence la manière dont l'ADN est compacté dans le noyau. De nouvelles expériences (tel que l'immunomarquage des histones ou la caractérisation protéique des deux formes de H3) sont nécessaires pour le vérifier.

Résultats clefs :

- ✓ Méthylome similaire à d'autres invertébrés avec méthylation.
- ✓ Structure chromatinienne inhabituelle ne permettant pas le ChIP sur histones.

VI.1.2. Effet des changements environnementaux sur l'épigénome

Dans l'objectif de mieux comprendre les possibles liens entre mécanismes épigénétiques et *Dauermodifikationen*, nous avons reproduit les changements environnementaux associés et observé leur impact sur l'épigénome, dans des conditions permettant (théoriquement) de maintenir le génotype constant. Chez *S. mansoni*, deux *Dauermodifikationen* sont documentées : le biais de sex-ratio lors du développement dans un hôte intermédiaire allopatrique (Lepesant *et al.* 2013) et l'induction de résistance au schistosomicide hycanthone (Jansma *et al.* 1977). La première correspond à un changement environnemental modéré et écologiquement réaliste, alors que la seconde est associée à un changement beaucoup plus brutal et artificiel (en milieu naturel, le parasite ne se trouve jamais en contact avec cette molécule). Chez *P. damicornis*, un phénomène de thermotolérance progressif et écologiquement réaliste a été mis en évidence sur le terrain et en laboratoire (Maynard *et al.* 2008; Vidal-Dupiol 2011; Putnam & Gates 2015).

VI.1.2.1. Résistance à l'hycanthone associée à des épimutations

Dans ses expériences sur l'hycanthone, Jansma *et al.* (1977) décrivent trois types d'induction de résistance. Nous nous sommes concentrés sur celle de type I. Elle se caractérise par l'injection chez des souris parasitées d'une dose d'hycanthone sous-curative pour *S. mansoni*. La progéniture des vers survivants présente alors une résistance importante à des doses curatives (*i.e.* tuant normalement >90% des parasites). Les conclusions de Jansma ont été assez débattues; certains auteurs arrivant à reproduire ses résultats d'induction suivant une héritabilité non-mendélienne (Brindley & Sher 1987;

Brindley *et al.* 1989) et d'autres décrivant plutôt un mécanisme de résistance monogénique récessif, déjà présent dans la population de schistosomes et sélectionné par le traitement à l'hycanthone (Dias & Olivier 1986; Pica-Mattoccia *et al.* 1991; Valentim *et al.* 2013). Face à ces divergences, nous sommes partis sans *a priori* et avons testé l'induction de type I sous une forme modifiée, inspirée par la célèbre expérience de Lederberg & Lederberg (1952). Cette expérience avait deux objectifs : **(I)** identifier l'origine préétablie ou induite de la résistance dans nos souches et **(II)** voir de quelle manière elle pouvait être liée à des modifications épigénétiques, par comparaison entre individus sensibles et résistants. Nous avons infecté des souris avec des cercaires, en théorie génétiquement identiques, issues de l'infection mono-miracidiale de mollusques. Si la résistance est présente ou absente chez le miracidium, tous les vers adultes issus de ce miracidium devraient répondre de manière similaire au traitement. Plutôt que de travailler sur la descendance comme Jansma, nous nous sommes intéressés directement aux adultes « induits » par le traitement. La raison pour cela était de n'avoir qu'un seul génotype (et donc sexe) du parasite dans l'hôte, pour faciliter la comparaison de l'épigénome en gardant un fond génétique constant, sans risque de contamination. Nous avons observé différentes réponses chez les vers adultes soumis au traitement d'induction à l'hycanthone. Au sein des vers adultes issus d'un même miracidium, bien que la majorité ait péri, nous avons obtenu des survivants. Cette observation penche vers une origine induite de la résistance, mais n'exclut pas la sélection des phénotypes dus à des mutations et/ou épimutations pendant la production des cercaires (discuté plus en profondeur dans la partie VI.2). En tout cas, nous n'avons pas retrouvé, dans les individus de la souche *SmBre* que nous avons utilisés, les variants génétiques associés à la résistance publiés par Valentim *et al.* (2013).

La comparaison des épigénomes des vers adultes contrôles (jamais en contact avec l'hycanthone) et des résistants a détecté 69 différences de structure chromatinienne, 43 d'entre elles situées au niveau de sites d'initiation de la transcription (TSS) de gènes. Une analyse d'ontologie génique a soulevé une surreprésentation des activités transférases, kinases et cataboliques parmi ces gènes, ce qui est consistant avec un processus de détoxification causé par l'élimination de l'hycanthone du métabolisme des vers résistants. Parmi ces gènes avec des épimutations, le *Multidrug Associated Resistance Protein 1* est un candidat intéressant, connu pour être associé à l'élimination de toxines métaboliques et de composés xénobiotiques chez d'autres espèces de vers trématodes.

À cause de l'arrêt de la commercialisation de l'hycanthone, nous n'avons pu tester les autres types d'induction décrits par Jansma *et al.* (1977) ou répéter nos expériences.

Résultats clefs :

- ✓ Des vers résistants à l'hycanthone apparaissent parmi des clones sensibles.
- ✓ Épimutations dans des gènes associés à la détoxification chez les résistants.

VI.1.2.2. Trois types d'épimutations lors des greffes de sporocystes

Dans la continuité des travaux de Julie Lepesant sur l'effet de l'hôte intermédiaire sur l'épigénome du *S. mansoni* (Lepesant 2012), nous avons employé la technique de greffe de sporocystes. Cette technique, si elle n'est pas écologiquement réaliste, permet de maintenir le génotype du parasite constant dans plusieurs hôtes intermédiaires mollusques, ce qui n'est pas possible en condition naturelle. Dans cette expérience, nous avons greffé deux différents génotypes femelles (Fa et Fb) du parasite (souche *SmBre*, Brésil) sur deux souches géographiquement distinctes du mollusque *Biomphalaria glabrata*. Une souche de *B. glabrata* est sympatrique du parasite (*BgBre*, Brésil), avec laquelle elle coévolve en milieu naturel, et l'autre est allopatrique (*BgGua*, Guadeloupe). Après la greffe de sporocystes, des cercaires ont été utilisées pour infecter des souris et obtenir des vers adultes. Cercaires et adultes ont été collectés et leur épigénome a été analysé par ChIP-Seq sur quatre modifications d'histones. Nous avons observé trois types d'épimutations chez les cercaires: **(I)** ciblées en réponse à l'environnement (51 épimutations), *i.e.* les différences sont communes à tous les individus d'une condition (allopatrique ou sympatrique), **(II)** associées au génotype (64 épimutations), *i.e.* des changements entre Fa ou Fb, quelle que soit la condition environnementale, et **(III)** aléatoires, *i.e.* sans lien avec le génotype ou la condition environnementale et différentes entre les réplicats (64 épimutations, dont 41 présentes dans au moins deux individus). Une majorité de ces changements se trouvent dans des régions transcrites : approximativement 77% pour les ciblés et reliés au génotype, et environ 80% pour les aléatoires. Les analyses d'ontologie géniques n'ont pas identifié de voies métaboliques ou d'enrichissement en fonctions moléculaires ou biologiques entre les gènes représentés dans chacun des trois types d'épimutations. L'analyse des séquences répétitives de *S. mansoni*, environ 48% du génome (Lepesant *et al.* 2012), n'a pas relevé de changement de la structure chromatinienne, ni entre conditions environnementales, ni entre génotypes ou clones.

Nous avons constaté que les modifications du nombre de copies (*copy number variations*, CNV) sont les seules sources de variations génétiques au sein des clones, et surviennent dans un ordre de grandeur similaire à celui des épimutations (entre 58 et 105 selon les clones observés). Aucune nouvelle insertion d'éléments transposables n'a été détectée entre les clones, ou entre les conditions environnementales sympatrique et allopatrique.

Résultats clefs :

- ✓ *Changements génétiques dans les CNV entre les clones et les conditions.*
- ✓ *Trois types d'épimutations : ciblées (51), liées au génotype (64), aléatoires (64).*
- ✓ *Pas de changements majeurs de la structure chromatinienne des régions répétées.*

VI.1.2.3. Altération du méthylome chez les coraux ayant subi un blanchissement

Des résultats préliminaires des comparaisons entre les trois colonies n'ayant jamais subi un blanchissement et les trois colonies ayant vécu deux épisodes de blanchissement ont relevé 2 898 régions différentiellement méthylées (DMR) communes à tous les réplicats. Ces régions d'un minimum de 300 paires de bases exhibent un minimum de 20% de différence en taux de cytosine méthylées. Même si ce résultat donne envie de faire un lien direct entre environnement, changement de méthylation et *Dauermodifikation*, il faut attendre la fin des analyses génétiques, transcriptomiques et de l'holobionte (clades de *Symbiodinium* dans les tissus) avant de tirer des conclusions.

Résultats clefs :

- ✓ *Environ 3 000 DMR entre coraux contrôles et ayant subi deux blanchissements.*

VI.1.3. Héritabilité des épimutations

Seule l'expérience de greffe de sporocyste chez *S. mansoni* a permis d'étudier l'héritabilité des modifications de la structure chromatinienne de manière somatique, entre deux stades développementaux du parasite (cercaires et adultes). Parmi les trois types d'épimutations identifiées, 76,6% de celles associées au génotype et 45,3% des aléatoires sont hérissables du stade cercaire au stade adulte. Aucune des épimutations ciblées par l'environnement et détectées chez les cercaires n'a été retrouvée chez les adultes. L'héritabilité à une génération suivante n'a pas pu être étudiée par cette expérience.

Néanmoins une seconde greffe de sporocystes a été faite pour des génotypes mâles et femelles (en conditions sympatrique et allopatrique), permettant d'obtenir une descendance sous forme de miracidia. Cercaires, adultes et miracidia de cette expérience ont été collectés et pourront être utilisés en ChIP-Seq pour poursuivre cette étude.

Chez *P. damicornis*, nos expériences ne permettent pas pour l'instant de déterminer s'il y a une héritabilité des régions différentiellement méthylées (DMR) entre coraux thermotolérants et naïfs (jamais soumis au stress thermique). La durée de vie des polypes au sein d'une colonie est inconnue. Il semble cependant possible de parler d'héritabilité somatique du phénotype de thermotolérance, car au fil des années, la colonie s'accroît en générant de nouveaux polypes (de manière asexuée) et ceux-ci résistent au blanchissement même s'ils n'ont eux-mêmes jamais vécu les épisodes de blanchissement précédents. Sur ce même raisonnement, comparer les méthylomes de nouvelles branches d'une année à l'autre pourrait permettre de répondre à cette interrogation.

Résultats clefs :

- ✓ *Épimutations aléatoires et associées au génotype sont hérissables somatiquement chez S. mansonii.*

VI.2. CONSIDÉRATIONS EXPÉRIMENTALES

Cette partie contient quelques réflexions sur la manière dont il faudrait aborder expérimentalement les futures études sur les *Dauermodifikationen*.

VI.2.1. Importance génétique et épigénétique de l'étude des *Dauermodifikationen*

Dans cette thèse, trois *Dauermodifikationen* ont été étudiées, et à chaque reprise des liens avec des modifications épigénétiques sont apparus. Si ces liens ne sont pour l'instant que corrélatifs, ils montrent cependant l'importance d'inclure les mécanismes épigénétiques dans l'investigation de phénotypes induits par l'environnement. Cet aspect est souvent négligé au seul profit de la génétique. L'étude de résistances à des composés xénobiotiques (pesticides, herbicides, antiparasitaires) en est un bon exemple. Seule une poignée d'articles suggère de possibles origines épigénétiques pour les résistances à des herbicides ou pesticides chez les plantes et animaux (May & Meyer 2003; Powles & Yu 2010; Gressel 2010; Cothran *et al.* 2013; Délye *et al.* 2013). Pourtant, il a été démontré à

plusieurs reprises, chez des cellules cancéreuses, que des mécanismes épigénétiques jouent un rôle déterminant dans la résistance aux traitements (Song *et al.* 2000; Roberti *et al.* 2006; Glasspool *et al.* 2006). On peut extrapoler que la même chose puisse survenir à l'échelle d'un tissu, organe ou individu. Quelques rares études montrent que ce phénomène existe à l'échelle d'un organisme entier. Chez *Myzus persicae*, un insecte ravageur des cultures, la résistance à plusieurs insecticides est sous contrôle épigénétique (Field & Blackman 2003). Nos travaux sur l'hycanthone vont dans le même sens, puisqu'ils décrivent une résistance induite associée à des changements de structure chromatinienne au sein de clones normalement sensibles (Roquis *et al.* 2013), en l'absence de la mutation génétique décrite par Valentim *et al.* (2013). D'autres recherches du même genre devraient bénéficier d'une analyse de la contribution épigénétique. Un bon exemple est celui de la résistance au praziquantel, la seule molécule actuellement préconisée pour le traitement de la schistosomiase (Webster *et al.* 2008). Comme pour l'hycanthone, la résistance au praziquantel est inductible par des doses sous-curatives. Elle est transmise de manière non-mendélienne et n'a pas d'origine génétique connue (Blanton *et al.* 2011; Greenberg 2013).

Les expériences présentées dans ce manuscrit ont pour la plupart été réalisées sur des clones. Ce point semblait essentiel pour déterminer quel était le rôle exact des mécanismes épigénétiques dans le phénomène de *Dauermodifikationen*. En utilisant des individus génétiquement distincts, même au sein d'une population de faible diversité génétique, il devient rapidement impossible de savoir quel est l'impact de chaque mutation ou épimutation sur le phénotype. Certaines séquences d'ADN ayant des effets en *cis* sur l'épigénome (Davidson 2010), cela complexifie encore plus l'observation de l'effet de l'environnement sur l'épigénome. Toutefois, même en utilisant des clones, il est essentiel de s'assurer que le patrimoine génétique reste véritablement constant entre les individus. Depuis les travaux de Barbara McClintock (McClintock 1953), on sait que les éléments transposables peuvent générer de nouveaux phénotypes, être rapidement mobilisés lors de stress (par exemple des changements environnementaux), et avoir un impact direct sur l'épigénome (Biemont 2010; Fedoroff 2012; Holoch & Moazed 2015). D'autres études ont mis en évidence une hétérogénéité génétique, non reliée à des éléments transposables, chez des individus ou populations cellulaires clonales de plusieurs espèces (Zhou *et al.* 2000; Bayne & Grevelding 2003; Saravanan & Godhe 2010; Li *et al.* 2015). Nos travaux sur *S. mansoni* ont montré pendant l'expérience de greffe de sporocystes qu'il y a des

changements de CNV entre sporocystes issus d'un même miracidium (cf. article soumis à Molecular Ecology, Chapitre IV). L'utilisation de clones réduit drastiquement la variabilité génétique entre individus, elle n'est pas une garantie absolue d'une absence totale de polymorphisme, et il faut s'assurer que le fond génétique reste bien constant entre les clones. Malgré cette importance, on trouve peu de publications étudiant les variations épigénétiques en réponse à l'environnement chez des espèces avec une descendance clonale, et il est très rare qu'une analyse génétique systématique sur l'ensemble du génome y soit incorporée. Quelques exemples sont visibles dans le **Tableau 2**.

Tableau 2 : Exemples de recherches sur les liens entre environnement, variabilité phénotypique et modifications épigénétiques sur des organismes clonaux (issus de multiplication asexuée).

Publication	Organisme utilisés	Variable environnementale	Modification épigénétique	Génétique vérifiée?
Asselman <i>et al.</i> (2015)	<i>Daphnia magna</i>	Multiples	Méthylation des cytosines	Non
Norouzitallab <i>et al.</i> (2014)	<i>Artemia spp.</i>	Choc thermique	Modifications d'histones	Non
Richards <i>et al.</i> (2012)	<i>Fallopia japonica</i>	Site de culture	Méthylation des cytosines	AFLP
Massicotte & Angers (2012)	<i>Chrosomus eos-neogaeus</i>	pH	Méthylation des cytosines	Non
Raj <i>et al.</i> (2011)	<i>Populus spp.</i>	Sècheresse	Méthylation des cytosines	Microsatellites
Gao <i>et al.</i> (2010)	<i>Alternanthera philoxeroides</i>	Site de culture	Méthylation des cytosines	AFLP
Verhoeven <i>et al.</i> (2009)	<i>Taraxacum officinale</i>	Herbivorie Pathogènes	Méthylation des cytosines	AFLP
Dombrovsky <i>et al.</i> (2009)	<i>Acyrtosiphon pisum</i>	Température Plante hôte	Méthylation des cytosines	Non
Field & Blackman (2003)	<i>Myzus persicae</i>	Insecticide	Méthylation des cytosines	Séquençage gène candidat

VI.2.2. Contributions respectives de l'épigénétique et la génétique au phénotype

Comme on vient de le voir, l'utilisation d'individus issus d'une multiplication clonale ne garantit pas d'avoir des individus génétiquement identiques. Dès lors, il n'est plus possible d'assigner avec certitude quels sont les rôles respectifs de la génétique et de l'épigénétique dans le phénotype. Des approches fonctionnelles, par inactivation de gène (knockout), mutations dirigées avec CRISPR-Cas9 (Jinek *et al.* 2012) ou édition épigénétique (Kofler *et al.* 2015) pourront éventuellement aider à identifier avec précision la source de la variabilité phénotypique. Toutefois, il faut garder en tête que génétique et épigénétique s'influencent intimement. On sait par exemple que l'insertion de nouveaux éléments transposables est susceptible de modifier les marques épigénétiques de la région où il s'insère, et que d'autres éléments régulateurs *cis* ont un impact direct sur l'épigénome (Biemont 2010; Holoch & Moazed 2015). De la même manière, des épimutations peuvent

influencer directement la séquence d'ADN. La 5-méthylcytosine peut être spontanément déaminée, la convertissant en thymine, ce qui est source de mutations ponctuelles (Duncan & Miller 1980). Selon les auteurs, la déamination de la 5mC surviendrait entre 20 à 40 fois plus souvent que n'importe quel autre type de mutations ponctuelles dans le génome (Russo *et al.* 2012). À une échelle génétique plus large, Skinner *et al.* (2015) ont soulevé que des épimutations étaient source de CNV dans les spermatozoïdes. Field & Blackman (2003) ont eux aussi associé des changements de méthylation de l'ADN avec une augmentation du nombre de copies d'un gène de résistance à un insecticide chez *Myzus persicae*. Similairement, nous avons observé l'apparition de CNV entre nos clones dans l'expérience de greffe de sporocystes (bien que nous n'ayons pas été en mesure de lui assigner une origine épigénétique directe), mais cela permet de présumer que ce phénomène est peut-être répandu. Ces réalités indiquent qu'il faut être prudent avant d'assigner une origine génétique ou épigénétique à un phénotype, et qu'il est temps de réfléchir à des approches plus intégratives de ces deux disciplines, au lieu d'essayer de les analyser séparément.

VI.2.3. Importance de l'analyse individuelle des réplicats biologique

Dans l'expérience de greffe de sporocystes, nous avons comparé trois individus entre une condition environnementale allopatrique et une sympatrique : deux clones du génotype Fb, et un du génotype Fa. Nous avons initialement effectué trois types de comparaisons chez les cercaires (et ensuite adultes) : **(I)** en utilisant tous les individus sympatriques vs allopatriques, pour identifier les épimutations ciblées par l'environnement **(II)** en comparant entre la condition allopatrique et sympatrique, un seul génotype à la fois, pour identifier des différences induites par l'environnement, mais seulement chez un des deux génotypes, et **(III)** en comparant des individus entre eux, peu importe la condition ou le génotype, pour détecter des épimutations aléatoires. Ce troisième point a été déterminant, car il a mis en évidence que les épimutations aléatoires sont présentes dans un ordre de grandeur similaire aux autres types (64 aléatoires pour 51 ciblées). Cette analyse s'est avérée particulièrement importante puisque nous avons ensuite constaté que 45% des épimutations aléatoires sont hérissables au stade adulte, alors qu'aucune ciblée ne l'est. Dans la littérature, je n'ai pas trouvé d'exemples d'analyses ayant utilisé cette approche un-à-un : soit il n'y a pas de réplicats (ce qui est de moins en moins fréquent), soit ces réplicats sont analysés comme un tout. Ce problème est aussi relié au fonctionnement

des rares logiciels permettant de faire des analyses comparatives entre deux ou plusieurs jeux de données de ChIP-Seq (Ernst & Kellis 2012; Schweikert *et al.* 2013; Shen *et al.* 2013; Chung *et al.* 2014; Elliott *et al.* 2014). La grande majorité nécessite impérativement des réplicats pour estimer la variabilité entre les échantillons et ensuite identifier les différences communes à tous les représentants d'une même condition. Lorsqu'une analyse un-à-un est possible, celle-ci reste toutefois contraignante, car il en sort énormément de faux-positifs (causé par le bruit de fond normalement éliminé lorsque l'on utilise des réplicats). Dans nos analyses sur la greffe de sporocyste, ce problème a été contourné en concentrant nos efforts sur les différences très marquées (puis par une inspection visuelle des résultats), au risque de perdre de la variabilité aléatoire réelle, mais moins marquée. Ne rechercher que les différences communes à tous les échantillons d'une condition peut mener à une interprétation biaisée. C'est pour cela qu'il est nécessaire de regarder l'ensemble de la variabilité épigénétique entre tous les individus. Nous allons maintenant généraliser ce genre d'approche à nos études épigénétiques, et l'appliquer à la comparaison des méthylomes de *P. damicornis* entre individus naïfs et thermotholérants.

VI.2.4. Mesure et quantification de la variabilité observable

En absence de méthodes d'épimutagénèse ciblées sur une large gamme d'organismes, il n'existe toujours pas de seuil biologique clair pour déterminer à quel moment une différence épigénétique entre deux conditions est biologiquement significative, et peut avoir un impact sur le phénotype. Dans nos analyses de ChIP-Seq, les différences d'intensités de pics observés entre deux échantillons sont parfois très fortes et flagrantes (présence ou absence de pics), et parfois plus subtiles (changement de la forme des pics, ou différence de taille moins marquée). Nous avons choisi une approche d'identification très rigoureuse, préférant les faux-négatifs aux faux-positifs. La même question peut se poser avec les analyses de méthylation de l'ADN. À partir de quel seuil de différence peut-on affirmer qu'une région est différentiellement méthylée par rapport à une autre? Empiriquement, nous avons choisi un seuil de 20% de différences en terme de cytosine méthylées, sur une longueur d'un minimum de 300 paires de bases, mais ces critères sont-ils véritablement réalistes? De plus, comme nous travaillons sur des organismes complets, il est plus facile de voir les différences présentes dans tous les tissus que les différences spécifiques à un tissu dont les signaux seront dilués par la présence de

tous les tissus. Les différences spécifiques à un tissu sont beaucoup plus difficiles à mettre en évidence dans des organismes de petite taille comme ceux utilisés durant ma thèse. Cependant on ne peut pas exclure qu'une différence très forte, spécifique à un tissu, soit détectable. Ainsi, on peut présumer que les fortes différences observées dans nos analyses sont, soit des changements dans une majorité des cellules de l'individu, soit des changements très forts dans un seul tissu. Dans ces deux cas, celles-ci sont vraisemblablement très significatives biologiquement.

VI.3. ORIGINE DE LA VARIABILITÉ ÉPIGÉNÉTIQUE

Les résultats de la greffe de sporocystes (Chapitre IV) ont montré différents niveaux de variabilité épigénétique : **(I)** ciblée par l'environnement, **(II)** associée à un génotype particulier, peut importe la condition environnementale et **(III)** aléatoire sans aucun lien entre le génotype ou la condition environnementale. Face à ces résultats, on peut réfléchir sur l'origine de la variabilité épigénétique observée. Commençons par les ciblées et aléatoires, pour lesquelles je peux envisager deux hypothèses principales.

Il pourrait s'agir d'un effet direct de l'environnement, ce dernier causant des réponses ciblées aidant l'organisme à mieux s'y adapter, et d'autres accidentelles (et aléatoires), n'ayant pas nécessairement un rôle adaptatif. On parlerait ici d'une origine externe de la variabilité épigénétique. Ici, l'environnement consiste en l'organisme du mollusque allopatrique ou sympatrique. Nous avons considéré une certaine homogénéité de cet environnement entre les mollusques d'une même souche, mais ce n'est peut-être pas forcément le cas. Ainsi, peut-être que les modifications aléatoires sont une réponse spécifique à l'individu mollusque dans lequel le sporocyste a été greffé. On aurait alors deux types de réponses ciblées par l'environnement : une première répondant à des caractéristiques communes à tous les mollusques d'une même souche, et une seconde plus spécifique au génotype du mollusque dans lequel le sporocyste a été greffé. Cependant, étant donné que 64% des épimutations aléatoires (41 sur 64) sont retrouvées chez au moins deux individus, peu importe la souche du mollusque ou le génotype du parasite, il semble peu probable qu'il s'agisse d'une réponse spécifique à l'individu mollusque. La fenêtre de temps durant laquelle surviendraient ces épimutations ciblées et aléatoires se situerait après la greffe, lors du développement dans le mollusque receveur.

L'autre possibilité est que la variabilité soit d'origine interne, c'est-à-dire qu'il existe des mécanismes permettant de créer une diversité épigénétique au sein d'un organisme, sans que cela ait un lien avec les conditions environnementales. La fenêtre de temps durant laquelle se produiraient ces épimutations est plus difficile à déterminer. Peut-être que cette variabilité aléatoire survient en tout temps et durant plusieurs stades développementaux du parasite, ou à des moments clefs, par exemple lors de la différenciation en sporocystes secondaires avant la greffe, ou bien encore en réponse à un stress. Les travaux de Remigi *et al.* (2014) et Marchetti *et al.* (2014) ont montré que certains mécanismes génétiques pouvaient causer une hypermutabilité dans le génome et accélérer l'évolution de la symbiose entre certaines bactéries du sol et des plantes légumineuses. Peut-être existe-t-il des mécanismes similaires provoquant une hyper-épi-mutabilité, permettant une meilleure adaptabilité de l'organisme. Si les épimutations sont préexistantes, l'environnement pourrait alors avoir un rôle de sélection. Les sporocystes possédant les épimutations les plus favorables (les ciblées) sont sélectionnés lors de la greffe et arrivent à survivre. Ceux n'ayant pas ces épimutations n'arrivent pas à s'implanter et meurent. Les épimutations aléatoires auraient alors un effet neutre, tout du moins au niveau du filtre de sélection environnemental appliqué. Il est finalement envisageable qu'il s'agisse d'une conséquence directe de la greffe. Cette technique n'est pas écologiquement réaliste, et peut-être que le développement particulier du sporocyste à ce moment-là (réversion du sp2 en sp1 puis la génération de nouveaux sp2) est une étape anormale générant une variabilité épigénétique aléatoire tout aussi anormale. Dans ce cas, la variabilité ciblée serait simplement un ajustement induit directement par l'environnement. J'exclus ici la possibilité que la source interne de la variabilité puisse se trouver dans les cercaires. Si cela avait été le cas, comme le ChIP-Seq est effectué sur des pools de près de 10 000 cercaires, une variabilité dans chaque cercaire aurait été noyée dans la population et ne serait pas apparue aussi clairement qu'elle ne l'est dans nos résultats.

Bien entendu, il est plausible qu'il y ait à la fois une origine interne et externe à cette variabilité épigénétique. L'héritabilité de ces variants peut nous suggérer que ce soit le cas. La différence d'héritabilité entre épimutations aléatoires, dont 45% sont transmises somatiquement au stade adulte, et épimutations ciblées par l'environnement, dont aucune n'est héritée, est particulièrement intrigante. Bien que cela reste spéculatif, il me semble concevable d'imaginer que les épimutations ciblées soient une réponse directe et flexible,

induite par l'environnement, pour permettre au sporocyste de s'acclimater à son environnement, sans que cela ne représente une contrainte évolutive ou un filtre de sélection. À l'inverse, la variabilité aléatoire serait d'origine interne, générant de la diversité phénotypique, et permettant le cas échéant de traverser des filtres de sélection environnementale. Cette idée est discutée en détail dans la partie VI.4.

La question sur l'origine induite (par l'environnement) ou aléatoire et préexistante de caractères, et leur sélection subséquente par l'environnement n'est pas nouvelle. Dans leur célèbre expérience, Lederberg & Lederberg (1952) ont utilisé des répliques sur membrane de velours pour transférer des bactéries cultivées sur un milieu non-sélectif vers des boîtes de Pétri avec un agent de sélection (antibiotique). Leurs résultats ont montré une origine spontanée du caractère, ce dernier étant ensuite sélectionné par l'environnement. Dans les expériences avec l'hycanthone (Chapitre III), l'injection du schistosomicide chez les souris infectées représente un stress artificiel, intense et rapide pour le parasite, la molécule agissant en l'espace de quelques heures (Pica-Mattoccia *et al.* 1988; 1992). La dose que nous avons utilisée était considérée comme sous-curative (60 mg/kg), mais tout de même assez importante par rapport à la dose thérapeutique (80 mg/kg) et provoquant la mort de nombreux parasites. Ici, il serait très surprenant que ce stress environnemental ait eu le temps d'induire des épimutations permettant une réponse du parasite. Il s'agit plutôt fort probablement de la sélection d'un phénotype préexistant chez le petit nombre d'adultes survivants. Ainsi, les épimutations que nous avons observées chez les vers résistants étaient déjà vraisemblablement présentes avant le stress. Dans notre expérience, comme dans celle des Lederbergs, le stress utilisé ne laisse pas beaucoup d'alternative : survie ou mort. Que se passerait-il avec un stress moins intense? Le développement des parasites dans des hôtes intermédiaires sympatriques et allopatriques est un stress plus faible, susceptible de survenir en milieu naturel, et nous observons un certain nombre de réponses ciblées entre tous les individus. Les analyses du méthylome de *P. damicornis* entre individus jamais stressés et thermotolérants pointent dans la même direction. L'intensité du stress environnemental pourrait moduler son impact sur l'épigénome de l'organisme : agent inducteur d'épimutations dans le cas de stress faibles et agent de sélection d'épimutations préexistantes dans le cas de stress plus violents. Une bonne manière de vérifier cette hypothèse serait d'adapter une fois de plus l'expérience de Lederberg, avec des individus génétiquement identiques (en théorie en tout cas), de les

exposer à différentes intensités d'un même stress (faible, fort, ou croissant) et d'observer l'impact sur l'épigénome. Cela aurait pu être fait, par exemple, en utilisant des doses beaucoup plus faibles d'hycanthone, causant une baisse de fitness mais sans létalité, et d'observer si on retrouvait les mêmes épimutations que celles identifiées dans les adultes survivants, mais de manière plus répandue au sein des clones.

Je termine sur le cas des épimutations associées au génotype. Initialement, leur origine me semblait très simple et très tranchée. Elle devait trouver leur source dans des mécanismes en *cis* par des séquences d'ADN ou d'autres mécanismes de maintenance spécifiques au génotype étudié. Toutefois, et encore une fois bien que cela reste très hypothétique, il pourrait s'agir d'épimutations aléatoires générées dans la précédente génération du parasite (*i.e.* les parents) et transmis ensuite à la descendance. Pour vérifier cette hypothèse, il faudrait suivre sur une ou deux générations du cycle de vie de *S. mansoni* comment évoluent ces épimutations aléatoires, et si celles transmises somatiquement entre cercaires et adultes finissent aussi par être héritées méiotiquement.

VI.4. RÔLE ÉVOLUTIF DE LA VARIABILITÉ ÉPIGÉNÉTIQUE

Victor Jollos définit les *Dauermodifikationen* comme des caractères acquis, transmis durant quelques générations successives, mais sans être incorporés dans l'hérédité permanente de la souche (Jollos 1924). À l'époque, la nature de cette héritabilité temporaire était inconnue, bien qu'il ait été suggéré qu'elle puisse être d'origine cytoplasmique. Dans les trois *Dauermodifikationen* présentées dans ce manuscrit (induction de résistance à l'hycanthone, effet de l'hôte intermédiaire sur le sex-ratio de *S. mansoni* et thermotolérance chez *P. damicornis*), nous avons pu faire corrélérer des mécanismes épigénétiques. Cela ne signifie pas qu'ils sont les seuls acteurs, mais ils jouent fort probablement un rôle significatif. Il est légitime de se demander quel est l'impact évolutif de ces *Dauermodifikationen*, et de la variabilité épigénétique qui s'y est trouvée associée.

L'hérédité des épimutations aléatoires lors de la greffe de sporocystes me semble être le résultat le plus intéressant sur cet aspect évolutif. Des cercaires génétiquement identiques montrent une variabilité épigénétique aléatoire, dont la moitié est transmise au stade développemental suivant. Cela signifie qu'il existe des mécanismes permettant de

générer une diversité épigénétique. En l'absence de données de transcriptomiques pour ces greffes, nous n'avons pas pu associer cette diversité à des phénotypes moléculaires et/ou adaptatifs, mais cela semble probable vu que 79% de ces épimutations se trouvent à l'intérieur de régions transcrites à un moment ou un autre du cycle de vie de *S. mansoni*. Pour l'instant, il n'est pas possible de savoir si ce mécanisme générateur d'épimutations est complètement aléatoire au sein du génome, ou s'il favorise des régions particulières. Cependant, l'absence de changements de structure chromatinienne au sein des séquences répétitives et la grande proportion d'épimutations dans des régions transcrites fait pencher la balance vers la seconde possibilité. L'héritabilité d'une proportion de ces épimutations fait en sorte que, si elles participent à un phénotype, elles peuvent contribuer à la sélection naturelle, et donc avoir une fonction évolutive.

On peut se poser la question de l'intérêt d'un autre système que celui de la mutation somatique pour générer de la diversité moléculaire et éventuellement phénotypique. La réponse est double. Premièrement, la réversibilité des épimutations offre un degré de plasticité qui n'est pas disponible avec la mutation somatique. En effet, dans l'éventualité d'une épimutation « défavorable », il y a toujours possibilité d'un retour en arrière, ce qui est moins facilement le cas avec la mutation. La seconde raison est liée à la fréquence d'apparition des épimutations. Les travaux de van der Graaf *et al.* (2015) ont montré que la vitesse d'apparition des épimutations est beaucoup plus élevée que celui des mutations ponctuelles. Ainsi, les épimutations offrent un moyen rapide et plastique de générer de la diversité, pouvant mener à de nouveaux phénotypes.

La prochaine étape sera de déterminer s'il existe une héritabilité transgénérationnelle des *Daeurmodifikationen* que nous avons étudiées. Nous soupçonnons que cela puisse être le cas. Pour le corail, les nouveaux polypes produits durant la croissance de la colonie sont thermotholérants même sans avoir connu de stress thermique. Les travaux de Jansma *et al.* (1977) et Brindley & Sher (1987) ont indiqué une héritabilité métastable de la résistance à l'hycanthone. Pour l'effet du développement du parasite dans des souches de mollusques différents, nous disposons déjà des échantillons de cercaires, adultes et miracidia d'une seconde expérience de greffe de sporocystes. Il sera relativement aisé de voir quelles sont les épimutations transmises lors de la reproduction sexuée. Une héritabilité à la génération suivante des épimutations aléatoires consoliderait leur importance évolutive.

Dans l'expérience de greffe de sporocystes, nous avons aussi observé des variations non-héritables, qu'elles soient ciblées par l'environnement ou aléatoires. Victor Jollos décrivait également des variations non transmises à la descendance lors de ses expériences sur les *Dauermodifikationen* (Jollos 1934). Darwin déclarait qu'une variation non-héritée n'a pas d'importance évolutive (Darwin 1859). Si on réfléchit à cette affirmation au sens évolutif strict, il est difficile de la contester. Cependant, n'est-il pas envisageable que des variations non-héritables puissent tout de même avoir un impact indirect sur l'évolution? Dans le cas des épimutations ciblées par l'environnement, si elles permettent une amélioration de la fitness lorsque l'individu se trouve dans un environnement sous-optimal, cela peut tout de même influencer la transmission des autres traits hérithables de l'individu. Je ferai ici un parallèle avec la notion de développement, intimement associée aux origines de l'épigénétique. Comme on l'a vu dans l'exploration de l'épigénome de *S. mansoni* (Chapitre II), le changement de stade développemental s'associe à des changements d'environnements très marqués (cercaires dans un milieu aquatique, et schistosomules/adultes dans l'hôte vertébré) et un bouleversement majeur et rapide (en l'espace de quelques heures) de la structure chromatinienne. La structure chromatinienne de chaque stade développemental n'est pas « héritée » d'un stade à l'autre, mais c'est sa répétition à chaque génération à un moment clef du développement qui permet au parasite de continuer son cycle de vie. Il en va peut-être de même pour les épimutations ciblées par nos changements d'hôtes allopatriques et sympatriques. Bien sûr, dans cet exemple comme dans celui du développement, on peut aussi considérer que c'est la plasticité de la structure chromatinienne qui est le trait hérithable et adaptatif, et pas les changements que nous observons directement.

Pour faire une synthèse de tous les résultats de mon travail sur les *Dauermodifikationen*, je propose un système à deux niveaux pour répondre aux changements de l'environnement. **(I)** Dans le cas d'un stress environnemental faible, ayant un impact léger sur la fitness, l'organisme peut répondre par des épimutations ciblées, offrant une plasticité à l'individu et lui permettant d'ajuster sa fitness. En l'absence du maintien du stress, ces changements ciblés ne persistent pas et ne sont pas hérités car le stress environnemental n'est pas suffisamment fort pour créer un filtre de sélection (**Figure 31**). Toutefois, si le stress est maintenu, on peut imaginer que ces épimutations persistent quelques générations, comme cela est suggéré dans les modèles de Pál & Miklós (1999) et

Klironomos *et al.* (2013). **(II)** Dans le cas d'un stress environnemental rapide et intense, il y aura une sélection des phénotypes les mieux adaptés. Ces phénotypes peuvent avoir comme source des mutations aléatoires, comme c'est classiquement admis, mais aussi des épimutations aléatoires. L'héritabilité de ces dernières permet de maintenir le phénotype sur plusieurs générations pour répondre efficacement au stress s'il se prolonge. Sur le long terme, ces épimutations peuvent se retrouver fixées (par conversion en mutation génétique par exemple) ou canalisées, si le stress perdure, ou être réversibles et régénérer le phénotype de départ ou de nouveaux phénotypes (**Figure 32**). Ce modèle reste à approfondir, mais me semble être en mesure d'expliquer les *Dauermodifikationen* observées par Victor Jollos ou le phénomène de canalisation décrit par Conrad Waddington.

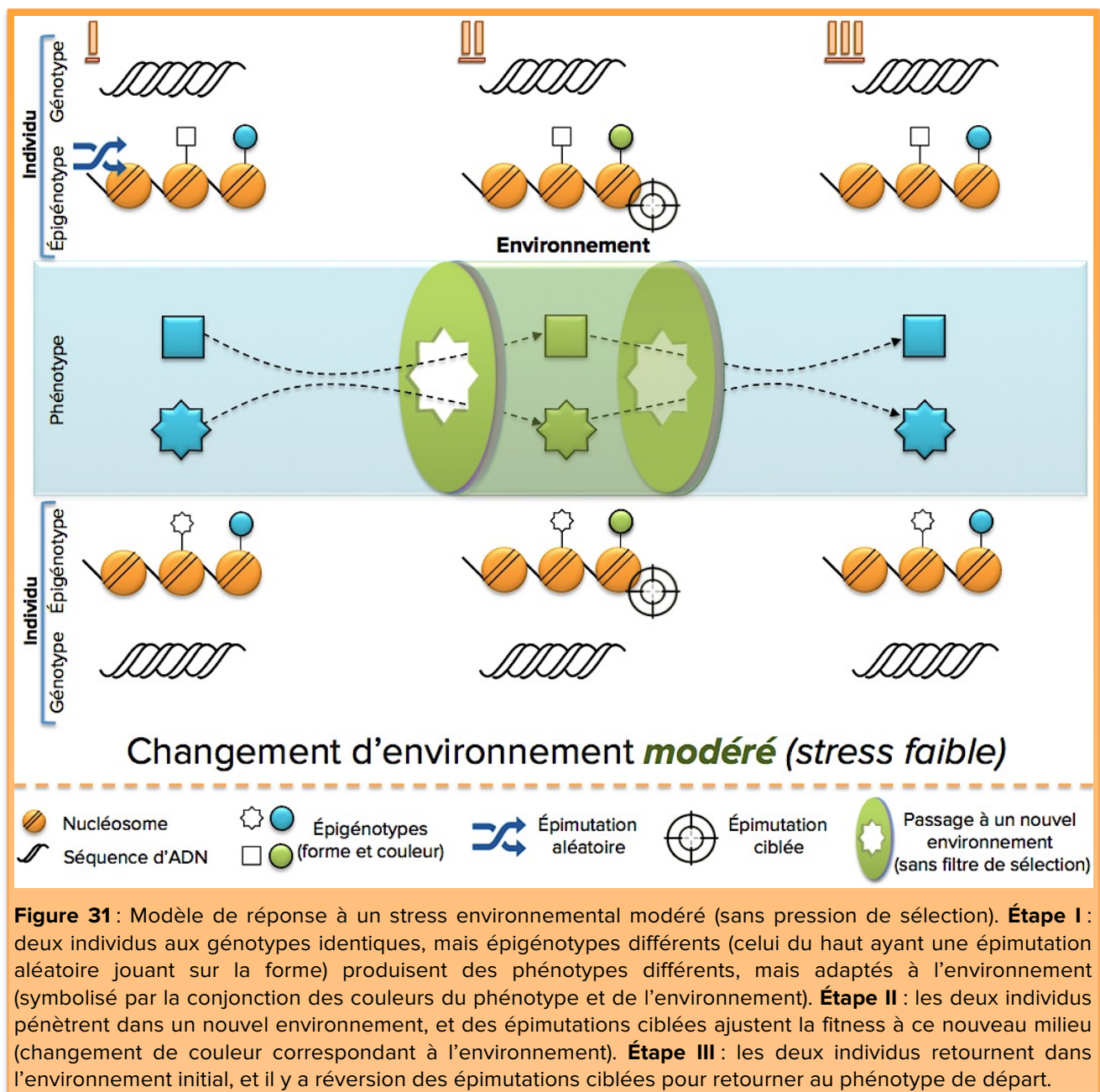


Figure 31 : Modèle de réponse à un stress environnemental modéré (sans pression de sélection). **Étape I :** deux individus aux génotypes identiques, mais épigénotypes différents (celui du haut ayant une épimutation aléatoire jouant sur la forme) produisent des phénotypes différents, mais adaptés à l'environnement (symbolisé par la conjonction des couleurs du phénotype et de l'environnement). **Étape II :** les deux individus pénètrent dans un nouvel environnement, et des épimutations ciblées ajustent la fitness à ce nouveau milieu (changement de couleur correspondant à l'environnement). **Étape III :** les deux individus retournent dans l'environnement initial, et il y a une réversion des épimutations ciblées pour retourner au phénotype de départ.

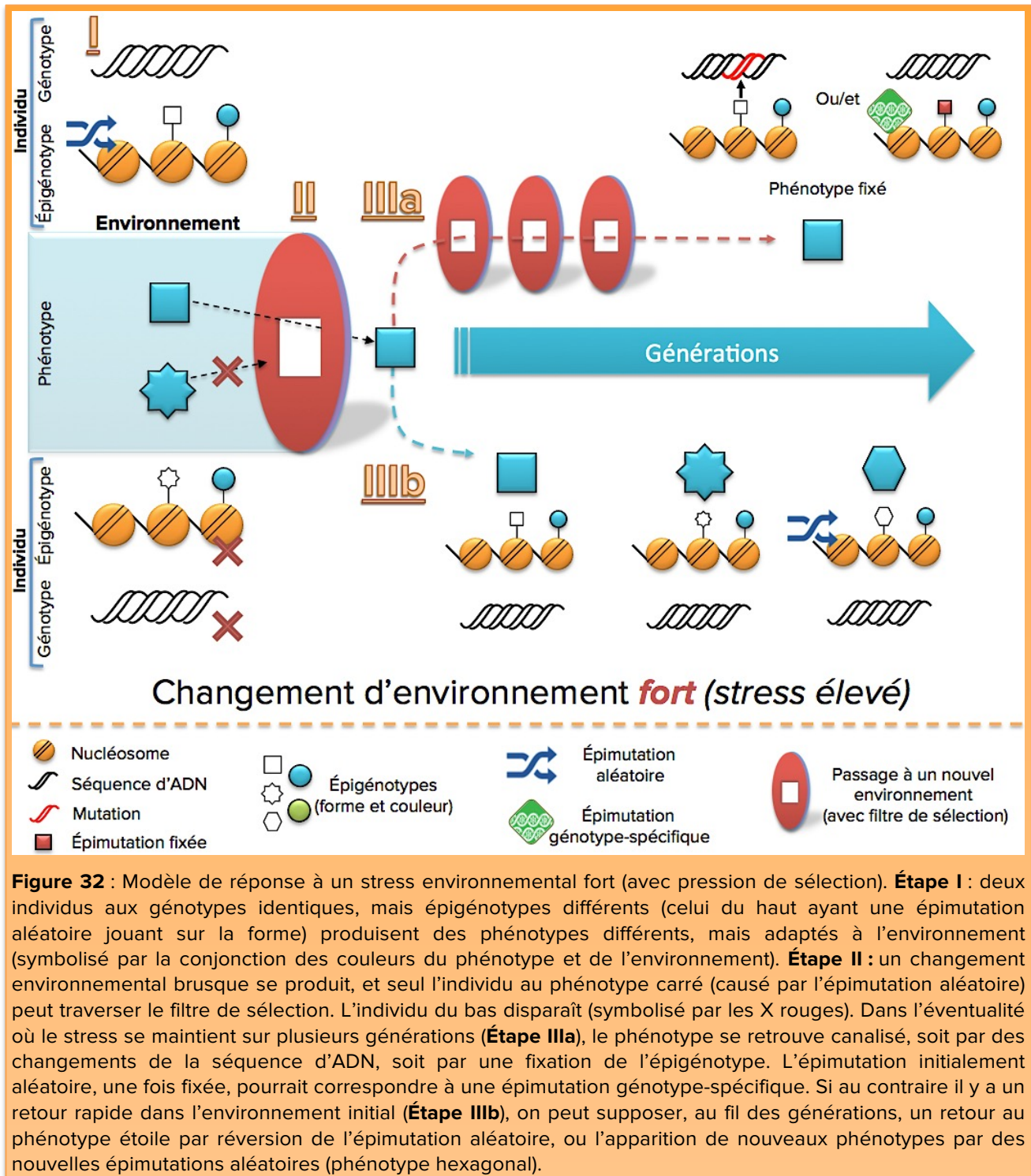


Figure 32 : Modèle de réponse à un stress environnemental fort (avec pression de sélection). **Étape I :** deux individus aux génotypes identiques, mais épigénotypes différents (celui du haut ayant une épimutation aléatoire jouant sur la forme) produisent des phénotypes différents, mais adaptés à l'environnement (symbolisé par la conjonction des couleurs du phénotype et de l'environnement). **Étape II :** un changement environnemental brusque se produit, et seul l'individu au phénotype carré (causé par l'épimutation aléatoire) peut traverser le filtre de sélection. L'individu du bas disparaît (symbolisé par les X rouges). Dans l'éventualité où le stress se maintient sur plusieurs générations (**Étape IIIa**), le phénotype se retrouve canalisé, soit par des changements de la séquence d'ADN, soit par une fixation de l'épigénotype. L'épimutation initialement aléatoire, une fois fixée, pourrait correspondre à une épimutation génotype-spécifique. Si au contraire il y a un retour rapide dans l'environnement initial (**Étape IIIb**), on peut supposer, au fil des générations, un retour au phénotype étoile par réversion de l'épimutation aléatoire, ou l'apparition de nouveaux phénotypes par des nouvelles épimutations aléatoires (phénotype hexagonal).

En me basant sur les expériences sur la résistance à l'induction et sur l'effet de la souche d'hôte intermédiaire (greffe de sporocyste), j'ai émis l'hypothèse qu'il puisse y avoir deux types de réponses épigénétique différentes en fonction de la nature du stress environnemental : **(I)** sélection d'épimutations préexistantes dans un stress fort, potentiellement létal, ou ayant un impact fortement négatif sur la fitness, et **(II)** induction d'épimutations par l'environnement dans un stress plus modéré.

Il serait intéressant d'observer sur plusieurs générations si l'intensité du stress et sa récurrence peuvent avoir un impact sur le taux d'épimutations aléatoires et/ou ciblées, et dans quelle mesure un phénotype causé par des épimutations ciblées peut être canalisé. Traditionnellement, on considère que des pressions de sélection fortes sont des moteurs de l'évolution adaptative, mais il est tout aussi envisageable que des pressions plus faibles soient en mesure d'avoir un rôle continu sur la dynamique évolutive, mais avec des effets (phénotypiques par exemples) plus subtils et moins évidents à détecter.

CHAPITRE VII

CONCLUSION ET PERSPECTIVES

Presque un siècle après la première description des *Dauermodifikationen* par Victor Jollos, nous commençons à avoir des éléments de réponses sur l'origine de ce phénomène. Durant cette thèse, j'ai cherché à savoir dans quelle mesure les modifications épigénétiques pouvaient être impliquées dans ces phénotypes induits par l'environnement, et à cerner leur héritabilité. Ces deux objectifs ont été atteints au travers des expériences présentées, mais il reste encore beaucoup à faire avant d'avoir le fin mot de l'histoire.

Sur le volet *Schistosoma*, une analyse plus pointue des différences de structure chromatinienne entre cercaires, schistosomules et adultes est en cours, et une publication est en préparation. Je propose ensuite de tenter de déterminer dans quelle mesure les épimutations observées sont hérissables méiotiquement. Les changements aléatoires de structures chromatiniennes hérités entre cercaires et adultes seront-ils transmis aux miracidia de la génération suivante? Les échantillons pour répondre à cette question sont déjà prêts, et permettront en plus d'approfondir nos connaissances sur l'épigénome des miracidia.

Sur le volet corail, nous savons maintenant qu'il ne faudra pas seulement regarder les régions différentiellement méthylées (DMR) communes à tous les réplicats d'une condition, mais également rechercher à caractériser la variabilité épigénétique aléatoire. Nous avons la chance d'avoir des données transcriptomiques, et il sera alors peut-être possible de corréler DMR et phénotypes moléculaires. Bien entendu, les analyses génétiques et de l'holobionte nous donneront aussi une vision plus claire de l'origine de la thermotolérance.

Globalement, ces travaux offrent aussi une réflexion sur les aspects expérimentaux à prendre en compte pour l'étude des mécanismes moléculaires sous-jacents aux *Dauermodifikationen* et pour mieux discriminer les implications de la génétique et de l'épigénétique dans les phénotypes induits par l'environnement. La prochaine étape sera d'ajouter une dimension fonctionnelle à nos analyses, de manière à ne plus rester dans le domaine du corrélatif. L'émergence des outils d'édition épigénétique, basé sur la protéine TAL, les doigts de zincs synthétiques et autres (Koflerle *et al.* 2015), devrait permettre l'accomplissement de cet objectif dans les années à venir.

CHAPITRE VIII

RÉFÉRENCES BIBLIOGRAPHIQUES

- A** Agrawal AA, Laforsch C, Tollrian R (1999) Transgenerational induction of defences in animals and plants. *Nature*, **401**, 60–63.
- Alleman M, Sidorenko L, McGinnis K *et al.* (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature*, **442**, 295–298.
- Arbuckle JH, Medveczky PG (2011) The molecular biology of human herpesvirus-6 latency and telomere integration. *Microbes and Infection*, **13**, 731–741.
- Asselman J, De Coninck DIM, Vandegehuchte MB *et al.* (2015) Global cytosine methylation in *Daphnia magna* depends on genotype, environment, and their interaction. *Environmental Toxicology and Chemistry*, **34**, 1056–1061.
- B** Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Nature Publishing Group*, **21**, 381–395.
- Barth TK, Imhof A (2010) Fast signals and slow marks: the dynamics of histone modifications. *Trends in Biochemical Sciences*, **35**, 618–626.
- Bayne CJ, Greveling CG (2003) Cloning of *Schistosoma mansoni* sporocysts in vitro and detection of genetic heterogeneity among individuals within clones. *Journal of Parasitology*, **89**, 1056–1060.
- Beltran S, Galinier R, Allienne JF, Boissier J (2008) Cheap, rapid and efficient DNA extraction method to perform multilocus microsatellite genotyping on all *Schistosoma mansoni* stages. *Memorias do Instituto Oswaldo Cruz*, **103**, 501–503.
- Berezney R (2002) Regulating the mammalian genome: the role of nuclear architecture. *Advances in enzyme regulation*, **42**, 39–52.
- Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature*, **447**, 407–412.
- Bernstein BE, Mikkelsen TS, Xie X *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315–326.
- Berriman M, Haas BJ, LoVerde PT *et al.* (2010) The genome of the blood fluke *Schistosoma mansoni*. *Nature*, **460**, 352–358.
- Biemont C (2010) A brief history of the status of transposable elements: from Junk DNA to major players in evolution. *Genetics*, **186**, 1085–1093.
- Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes & Development*, **16**, 6–21.
- Bird A (2007) Perceptions of epigenetics. *Nature*, **447**, 396–398.
- Blanton RE, Blank WA, Costa JM *et al.* (2011) *Schistosoma mansoni* population structure and persistence after praziquantel treatment in two villages of Bahia, Brazil. *International journal for parasitology*, **41**, 1093–1099.
- Bogitsh BJ, Carter CE, Oeltmann TN (2013) *Human Parasitology*. Academic Press.
- Botcheva K, McCorkle SR, McCombie WR, Dunn JJ, Anderson CW (2011) Distinct p53 genomic binding patterns in normal and cancer-derived human cells. *Cell Cycle*, **10**, 4237–4249.

- B** Bourne G, Moir C, Bikkul U *et al.* (2013) Interphase chromosome behavior in normal and diseased cells. In: pp. 9–33. Springer New York, New York, NY.
- Bowler PJ (2003) *Evolution: the history of an idea*. University of California Press.
- Boyle S, Rodesch MJ, Halvensleben HA, Jeddloh JA, Bickmore WA (2011) Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Research*, **19**, 901–909.
- Brindley PJ, Sher A (1987) Anti-schistosomal drugs: observations on the mechanism of drug resistance to hycanthone, and on the involvement of host antibodies in the mode of action of praziquantel. *Memorias do Instituto Oswaldo Cruz*, **82 Suppl 4**, 157–161.
- Brindley PJ, Lewis FA, McCutchan TF, Bueding E, Sher A (1989) A genomic change associated with the development of resistance to hycanthone in *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, **36**, 243–252.
- Brink RA (1941) Victor Jollos 1887-1941. *Science (New York, N.Y.)*, **94**, 270–272.
- Bygren LO, g PT, Carstensen J *et al.* (2014) Change in paternal grandmothers' early food supply influenced cardiovascular mortality of the female grandchildren. *BMC Genetics*, **15**, 1–6.
- C** Cabezas-Cruz A, Lancelot J, Caby S, Oliveira G, Pierce RJ (2014) Epigenetic control of gene function in schistosomes: a source of therapeutic targets? *Frontiers in genetics*, **5**, 1–12.
- Cai P, Piao X, Hao L *et al.* (2013) A deep analysis of the small non-coding RNA population in *Schistosoma japonicum* Eggs (E Dias-Neto, Ed.). *PLoS ONE*, **8**, e64003.
- Carroll WE (2000) Creation, evolution and Thomas Aquinas. *Revue des Questions Scientifiques*, **171**, 319–348.
- Cech TR, Steitz JA (2014) The noncoding RNA revolution— trashing old rules to forge new ones. *Cell*, **157**, 77–94.
- Chahwan R, Wontakal SN, Roa S (2010) Crosstalk between genetic and epigenetic information through cytosine deamination. *Trends in Genetics*, **26**, 443–448.
- Chan SWL, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Reviews Genetics*, **6**, 351–360.
- Chung D, Zhang Q, Keleş S (2014) MOSAiCS-HMM: A model-based approach for detecting regions of histone modifications from ChIP-seq data. In: *Statistical Analysis of Next Generation Sequencing Data* 277–295. Springer.
- Cioli D, Pica-Mattoccia L, Moroni R (1992) *Schistosoma mansoni*: hycanthone/oxamniquine resistance is controlled by a single autosomal recessive gene. *Experimental parasitology*, **75**, 425–432.
- Clément JAJ, Toulza E, Gautier M *et al.* (2013) Private selective sweeps identified from next-generation pool-sequencing reveal convergent pathways under selection in two inbred *Schistosoma mansoni* strains (AR Jex, Ed.). *PLoS Neglected Tropical Diseases*, **7**, e2591.
- Cooper TF, Lai M, Ulstrup KE *et al.* (2011) *Symbiodinium* genotypic and environmental controls on lipids in reef building corals. *PLoS ONE*, **6**, e20434.

- C** Cosseau C, Grunau C (2011) Native chromatin immunoprecipitation. *Methods in molecular biology* (Clifton, N.J.), **791**, 195–212.
- Cosseau C, Azzi A, Rognon A *et al.* (2010) Epigenetic and phenotypic variability in populations of *Schistosoma mansoni* - a possible kick-off for adaptive host/parasite evolution. *Oikos*, **119**, 669–678.
- Cosseau C, Azzi A, Smith K *et al.* (2009) Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters. *Molecular and Biochemical Parasitology*, **166**, 70–76.
- Cothran RD, Brown JM, Relyea RA (2013) Proximity to agriculture is correlated with pesticide tolerance: evidence for the evolution of amphibian resistance to modern pesticides. *Evolutionary Applications*, **6**, 832–841.
- Counce SJ, Sander K, Fässler PE *et al.* (2012) *Landmarks in developmental biology 1883–1924: historical essays from Roux’s archives*. Springer Berlin Heidelberg.
- Coura JR, Wanke B, Figueiredo N (1974) Evolutive pattern of schistosomiasis and life-span of *S. mansoni* in patients living in non-endemic area in Brazil. *Rev Soc Bras Med*, **3**, 193–198
- Couto FFBF, Coelho PMZP, Araújo NN *et al.* (2011) *Schistosoma mansoni*: a method for inducing resistance to praziquantel using infected *Biomphalaria glabrata* snails. *Memorias do Instituto Oswaldo Cruz*, **106**, 153–157.
- Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*, **2**, 292–301.
- Cremer T, Cremer M (2010) Chromosome Territories. *Cold Spring Harbor Perspectives in Biology*, **2**, a003889–a003889.
- Creyghton MP, Cheng AW, Welstead GG *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, **107**, 21931–21936.
- D** da Rocha Pitta MG, da Rocha Pitta MG, de Melo Rêgo MJB, Galdino SL (2013) The evolution of drugs on *Schistosoma* treatment: looking to the past to improve the future. *Mini reviews in medicinal chemistry*, **13**, 493–508.
- Danchin E (2013) Avatars of information: towards an inclusive evolutionary synthesis. *Trends in Ecology & Evolution*, **28**, 351–358.
- Danchin E, Charmantier A, Champagne FA *et al.* (2011) Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nature Reviews Genetics*, **12**, 475–486.
- Darwin C (1859) *On the origins of species by means of natural selection*. London Murray.
- Darwin C (1871) Pangenesis. *Nature*, **3**, 502–503.
- Darwin C (1872) *The descent of Man, and selection in relation to sex*. D. Appleton.
- Darwin C (1892) *The variations of animals and plants under domestication*. D. Appleton.
- Davidson EH (2010) *The regulatory genome: gene regulatory networks in development and evolution*. Elsevier Science.

- D Dawkins R (2006) *The Selfish Gene : 30th Anniversary edition*. OUP Oxford.
- de Lamarck JB (1809) *Philosophie zoologique, ou exposition des considérations relatives à l'histoire naturelle des animaux*. Lanoe.
- de Lamarck J-B de M (1815) *Histoire naturelle des animaux sans vertèbres, volume I*. Verdière.
- de Souza Gomes M, Muniyappa MK, Carvalho SG, Guerra-Sá R, Spillane C (2011) Genome-wide identification of novel microRNAs and their target genes in the human parasite *Schistosoma mansoni*. *Genomics*, **98**, 96–111.
- Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. *Genes & Development*, **25**, 1010–1022.
- DeJong RJ, Morgan JA, Paraense WL et al. (2001) Evolutionary relationships and biogeography of *Biomphalaria* (Gastropoda: Planorbidae) with implications regarding its role as host of the human bloodfluke, *Schistosoma mansoni*. *Molecular Biology and Evolution*, **18**, 2225–2239.
- Délye C, Jasieniuk M, Le Corre V (2013) Deciphering the evolution of herbicide resistance in weeds. *Trends in Genetics*, **9**, 649–658.
- Dias LC de S, Olivier CE (1986) Failure at inducing resistance to schistosomicidal drugs in a Brazilian human strain of *Schistosoma mansoni*. *Revista do Instituto de Medicina Tropical de São Paulo*, **28**, 352–357.
- Dias LC de S, Bruce JI, Coles GC (1988a) Strain variation in the infectivity of *Schistosoma mansoni* for *Biomphalaria glabrata*. *Revista do Instituto de Medicina Tropical de São Paulo*, **30**, 86–90.
- Dias LC, Bruce JI, Coles GC (1988b) Variation in response of *Schistosoma mansoni* strains to schistosomicides. *Revista do Instituto de Medicina Tropical de São Paulo*, **30**, 81–85.
- Dietrich MR (1994) The origins of the neutral theory of molecular evolution. *Journal of the history of biology*, **27**, 21–59.
- Dietrich MR (1996) On the mutability of genes and geneticists: The “Americanization” of Richard Goldschmidt and Victor Jollos. *Perspectives on Science*, **4**, 321–345.
- Dobzhansky T, Dobzhansky TG (1970) *Genetics of the Evolutionary Process*. Columbia University Press.
- Dombrovsky A, Arthaud L, Ledger TN, Tares S, Robichon A (2009) Profiling the repertoire of phenotypes influenced by environmental cues that occur during asexual reproduction. *Genome Research*, **19**, 2052–2063.
- Du J, Johnson LM, Jacobsen SE, Patel DJ (2015) DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews Molecular Cell Biology*, **16**, 519–532.
- Dunbar HE, Wilson A, Ferguson NR, Moran NA (2007) Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biology*, **5**, 1006–1015.
- Duncan BK, Miller JH (1980) Mutagenic deamination of cytosine residues in DNA. *Nature*, **287**, 560–561.

- E** Ebert D, Hamilton WD (1996) Sex against virulence: the coevolution of parasitic diseases. *Trends in Ecology & Evolution*, **11**, 79-82.
- Elliott HR, Tillin T, McArdle WL et al. (2014) Differences in smoking associated DNA methylation patterns in South Asians and Europeans. **6**, 1–10.
- Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods*, **9**, 215–216.
- F** Fan T-Y, Li J-J, le S-X, Fang L-S (2002) Lunar periodicity of larval release and larval development of *Pocillopora damicornis* in Thailand. *Zoological Studies*, **41**, 288–294.
- Fantappiè RM, Gimba RPE, Rumjanek FD (2001) Lack of DNA methylation in *Schistosoma mansoni*. *Experimental parasitology*, **98**, 162–166.
- Fedoroff NV (2012) Transposable elements, epigenetics, and genome evolution. *Science*, **338**, 758–767.
- Feil R, Fraga MF (2012) Epigenetics and the environment: emerging patterns and implications. *Nature Reviews Genetics*, **13**, 97–109.
- Felsenfeld G (2014) A brief history of epigenetics. *Cold Spring Harbor Perspectives in Biology*, **6**, a018200–a018200.
- Ferrari MLA, Coelho PMZ, Antunes CMF, Tavares CAP, da Cunha AS (2003) Efficacy of oxamniquine and praziquantel in the treatment of *Schistosoma mansoni* infection: a controlled trial. *Bulletin of The World Health Organization*, **81**, 190–196.
- Field LM, Blackman RL (2003) Insecticide resistance in the aphid *Myzus persicae* (Sulzer): chromosome location and epigenetic effects on esterase gene expression in clonal lineages. *Biological Journal of the Linnean Society*, **79**, 107-113.
- Filion GJ, van Bemmelen JG, Braunschweig U et al. (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*, **143**, 212–224.
- Fisher RA (1930) *The genetical theory of natural selection*. At the Clarendon Press.
- Flores KB, Amdam GV (2011) Deciphering a methylome: what can we read into patterns of DNA methylation? *Journal of Experimental Biology*, **214**, 3155–3163.
- Flores KB, Wolschin F, Amdam GV (2013) The role of methylation of DNA in environmental adaptation. *Integrative and Comparative Biology*, **53**, 359–372.
- Fneich S (2014) La contribution de l'épigénétique dans l'expression des variants phénotypiques essentiels pour l'interaction: *Schistosoma mansoni* / *Biomphalaria glabrata*. Université de Perpignan Via Domitia, Perpignan.
- Fneich S, Dheilly N, Adema C et al. (2013) 5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of *Biomphalaria glabrata*, a snail intermediate host of *Schistosoma mansoni*. *Parasites & Vectors*, **6**, 1–1.
- Foster HA, Griffin DK, Bridger JM (2012) Interphase chromosome positioning in in vitro porcine cells and ex vivo porcine tissues. *BMC cell biology*, **13**, 30.
- Fu Y, Luo G-Z, Chen K et al. (2015) N6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell*, **161**, 1–14.
- Futuyma DJ (2013) *Natural selection and adaptation*. The Princeton Guide to Evolution.

- G Gao L, Geng Y, Li B, Chen J, Yang J (2010) Genome-wide DNA methylation alterations of *Alternanthera philoxeroides* in natural and manipulated habitats: implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. *Plant, Cell & Environment*, **33**, 1820–1827.
- Gavery MR, Roberts SB (2010) DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). *BMC Genomics*, **11**, 483.
- Geyer KK, Chalmers IW, MacKintosh N *et al.* (2013) Cytosine methylation is a conserved epigenetic feature found throughout the phylum Platyhelminthes. *BMC Genomics*, **14**, 462.
- Geyer KK, pez CMRIGLO, Chalmers IW *et al.* (2011) Cytosine methylation regulates oviposition in the pathogenic blood fluke *Schistosoma mansoni*. *Nature Communications*, **2**, 424–10.
- Gibney ER, Nolan CM (2010) Epigenetics and gene expression. *Heredity*, **105**, 4–13.
- Gilbert SF (2000) *Developmental Biology*. Palgrave Macmillian.
- Glasspool RM, Teodoridis JM, Brown R (2006) Epigenetics as a mechanism driving polygenic clinical drug resistance. *British Journal of Cancer*, **94**, 1087–1092.
- Goecks J, Nekrutenko A, Taylor J, Galaxy Team T (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, **11**, R86.
- Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: a landscape takes shape. *Cell*, **128**, 635–638.
- Goll MG, Kirpekar F, Maggert KA *et al.* (2006) Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science (New York, N.Y.)*, **311**, 395–398.
- Goodrum MR (2002) Atomism, atheism, and the spontaneous generation of human beings: the debate over the natural origin of the first humans in seventeenth-century Britain. *Journal of the history of ideas*, **63**, 207–224.
- Gordin MD (2011) How lysenkoism became pseudoscience: Dobzhansky to Velikovsky. *Journal of the history of biology*, **45**, 443–468.
- Greenberg RM (2013) New approaches for understanding mechanisms of drug resistance in schistosomes. *Parasitology*, **140**, 1534–1546.
- Greer EL, Blanco MA, Gu L *et al.* (2015) DNA methylation on N6-adenine in *C. elegans*. *Cell*, **161**, 1–11.
- Gressel J (2010) Low pesticide rates may hasten the evolution of resistance by increasing mutation frequencies. *Pest Management Science*, **67**, 253–257.
- Gryseels B, Mbaye A, De Vlas SJ *et al.* (2001) Are poor responses to praziquantel for the treatment of *Schistosoma mansoni* infections in Senegal due to resistance? An overview of the evidence. *Tropical medicine & international health : TM & IH*, **6**, 864–873.
- Gryseels B, Polman K, Clerinx J, Kestens L (2006) Human schistosomiasis. *The Lancet*.
- Gu SG, Fire A (2009) Partitioning the *C. elegans* genome by nucleosome modification, occupancy, and positioning. *Chromosoma*, **119**, 73–87.

- G** Guerrero-Bosagna C, Skinner MK (2011) Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. *Molecular and Cellular Endocrinology*, **354**, 1–6.
- Guillou F, Roger E, Moné Y *et al.* (2007) Excretory–secretory proteome of larval *Schistosoma mansoni* and *Echinostoma caproni*, two parasites of *Biomphalaria glabrata*. *Molecular and Biochemical Parasitology*, **155**, 45–56.
- Gutierrez A (2004) Evolution of dnmt-2 and mbd-2-like genes in the free-living nematodes *Pristionchus pacificus*, *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Nucleic Acids Research*, **32**, 6388–6396.
- H** Hackett JD, Scheetz TE, Yoon HS *et al.* (2005) Insight into a dinoflagellate genome through expressed sequence tag analysis. *BMC Genomics*, **6**, 80.
- Hahnen E, Hauke J, Tränkle C *et al.* (2008) Histone deacetylase inhibitors: possible implications for neurodegenerative disorders. *Expert Opinion on Investigational Drugs*, **17**, 169–184.
- Hall BK, Olson W (2006) Keywords and concepts in evolutionary developmental biology. *Harvard University Press*.
- Hamilton WD (1964) The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, **7**, 1–16.
- Harman O (2011) *The Price of Altruism: George Price and the Search for the Origins of Kindness*. W. W. Norton.
- Harrison PL (2011) Sexual reproduction of scleractinian corals. In: *Coral Reefs: An Ecosystem in Transition*, 59–85. Springer Science & Business Media.
- He X-J, Chen T, Zhu J-K (2011) Regulation and function of DNA methylation in plants and animals. *Nature Publishing Group*, **21**, 442–465.
- Hewish DR, Burgoyne LA (1973) Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochemical and biophysical research communications*, **52**, 504–510.
- Heyward AJ, Negri AP (2012) Turbulence, cleavage, and the naked embryo: a case for coral clones. *Science (New York, N.Y.)*, **335**, 1064–1064.
- Hoegh-Guldberg O (2011) Coral reef ecosystems and anthropogenic climate change. *Regional Environmental Change*, **11**, 215–227.
- Hoegh-Guldberg O, Mumby PJ, Hooten AJ *et al.* (2007) Coral Reefs Under Rapid Climate Change and Ocean Acidification. *Science*, **318**, 1737–1742.
- Holeski LM, Jander G, Agrawal AA (2012) Transgenerational defense induction and epigenetic inheritance in plants. *Trends in Ecology & Evolution*, **27**, 618–626.
- Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. *Science*, **187**, 226–232.
- Holoch D, Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. *Nature Reviews Genetics*, **16**, 71–84.

- H Huang Y, Rao A (2014) Connections between TET proteins and aberrant DNA modification in cancer. *Trends in Genetics*, **30**, 664-674.
- Hughes RN, Hughes DJ, Smith IP, Dale AC (2015) *Oceanography and marine biology: an annual review*. CRC Press.
- Hughes TP, Bellwood DR, Connolly SR (2002) Biodiversity hotspots, centres of endemism, and the conservation of coral reefs. *Ecology Letters*, **5**, 775–784.
- Hunt BG, Brisson JA, Yi SV, Goodisman MAD (2010) Functional conservation of DNA methylation in the pea aphid and the honeybee. *Genome Biology and Evolution*, **2**, 719-728.
- Hurd PJ (2011) The era of epigenetics. *Briefings in Functional Genomics*, **9**, 425–428.
- Huxley J (1943) *Evolution: the modern synthesis*. Harper & Brothers Publishers.
- I Ismail MM, Taha SA, Farghaly AM, el-Azony AS (1994) Laboratory induced resistance to praziquantel in experimental schistosomiasis. *Journal of the Egyptian Society of Parasitology*, **24**, 685–695.
- IUCN *Pocillopora damicornis* (2014): Hoeksema BW, Rogers A & Quibilan MC. Available at [http:// www.iucnredlist.org/details/133222/0](http://www.iucnredlist.org/details/133222/0).
- J Jablonka E, Lamb MJ (2002) The changing concept of epigenetics. *Annals of the New York Academy of Sciences*, **981**, 82–96.
- Jablonka E, Lamb MJ (2005) *Evolution in Four Dimensions*. MIT Press.
- Jablonka E, Lamb MJ (2008) Soft inheritance: challenging the modern synthesis. *Genetics and Molecular Biology*, **31**, 389–395.
- Jablonka E, Lachmann M, Lamb MJ (1992) Evidence, mechanisms and models for the inheritance of acquired characters. *Journal of Theoretical Biology*, **158**, 245–268.
- Jamison DT, Breman JG, Measham AR et al. (2006) *Helminth infections: soil-transmitted helminth infections and schistosomiasis*. World Bank, Washington (DC).
- Jansma WBW, Rogers SHS, Liu CLC, Bueding EE (1977) Experimentally produced resistance of *Schistosoma mansoni* to hycanthone. *The American journal of tropical medicine and hygiene*, **26**, 926–936.
- Jensen P (2013) Transgenerational epigenetic effects on animal behaviour. *Progress in Biophysics and Molecular Biology*, **113**, 447-454.
- Jensen T, Ronne-Rasmussen JO, Bygbjerg IC (1995) Acute schistosomiasis (Katayama fever). *Ugeskrift for laeger*, **157**, 4825–4827.
- Jeon J, Choi J, Lee G-W et al. (2015) Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae*. *Scientific Reports*, **5**, 8567.
- Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews Genetics*, **10**, 161–172.

- J Jinek M, Chylinski K, Fonfara I *et al.* (2012) A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (New York, N.Y.)*, **337**, 816–821.
- Johannes F, Colome-Tatche M (2011) Quantitative epigenetics through epigenomic perturbation of isogenic lines. *Genetics*, **188**, 215–227.
- Johannsen W, Selskab DKDV (1903) *Om arvelighed i samfund og i rene linier*. Det kongelige Danske Videnskabernes Selskab.
- Johannsen W (1905) *Arvelighedslærens elementer: forelæsninger holdte ved Københavns universitet*. Gyldendal.
- Johannsen W (1909) Elemente der exakten Erblchkeitslehre. *Molecular and General Genetics MGG*, **2**, 136–137.
- Jollos V (1921) *Untersuchungen über Variabilität und Vererbung bei Infusorien*. Fischer.
- Jollos V (1924) *Experimentelle Prostitenstudien. I. Untersuchungen über Variabilität und Vererbung bei Infusorien*. Molecular and General Genetics MGG.
- Jollos V (1934) Inherited changes produced by heat-treatment in *Drosophila melanogaster*. *Genetica*, **16**, 476–494.
- Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, **13**, 484–492.
- Jourdane J (1984) Maintenance of a male and a female clone of *Schistosoma mansoni* by microsurgical transplantation of sporocysts. Reliability of the method. *Annales de parasitologie humaine et comparée*, **59**, 361–367.
- Jourdane J, Theron A (1980) *Schistosoma mansoni*: cloning by microsurgical transplantation of sporocysts. *Experimental parasitology*, **50**, 349–357.
- K Kaati G, Bygren LO, Edvinsson S (2002) Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *European Journal of Human Genetics*, **10**, 682–688.
- Kaati G, Bygren LO, Pembrey M, Sjöström M (2007) Transgenerational response to nutrition, early life circumstances and longevity. *European Journal of Human Genetics*, **15**, 784–790.
- Kharchenko PV, Alekseyenko AA, Schwartz YB *et al.* (2010) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature*, **471**, 480–485.
- King CH (2010) Parasites and poverty: The case of schistosomiasis. *Acta Tropica*, **113**, 95–104.
- Kirk GS, Raven JE, Schofield M (1983) *The presocratic philosophers: a critical history with a selection of texts*. Cambridge University Press.
- Köferle A, Stricker SH, Beck S (2015) Brave new epigenomes: the dawn of epigenetic engineering. *Genome Medicine*, **7**, 1–3.
- Klironomos FD, Berg J, Collins S (2013) How epigenetic mutations can affect genetic evolution: Model and mechanism. **35**, 571–578.
- Kramarsky-Winter E, Fine M, Loya Y (1997) Coral polyp expulsion. *Nature*, **387**, 137.

- L** Laforsch C, Tollrian R (2004) Inducible defenses in multipredator environments: cyclomorphosis in *Daphnia cucullata*. *Ecology*, **85**, 2302–2311.
- Laskey RA, Gurdon JB (1970) Genetic content of adult somatic cells tested by nuclear transplantation from cultured cells. *Nature*, **228**, 1332–1334.
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, **11**, 204–220.
- Lederberg J, Lederberg EM (1952) Replica plating and indirect selection of bacterial mutants. *Journal of bacteriology*, **63**, 399–406.
- Lefevre P, Bonifer C (2006) Analyzing histone modification using crosslinked chromatin treated with micrococcal nuclease. *Methods in molecular biology*, **325**, 315–325.
- Lepesant J (2012) Étude de la variabilité phénotypique, sous influence environnementale, chez *Schistosoma mansoni*, parasite de l'homme : implication de l'épigénome par une approche globale. Université de Perpignan Via Domitia.
- Lepesant MJM, Boissier J, Climent D, Cosseau C, Grunau C (2013) Female biased sex-ratio in *Schistosoma mansoni* after exposure to an allopatric intermediate host strain of *Biomphalaria glabrata*. *Experimental parasitology*, **135**, 350–356.
- Lepesant MJM, Roquis D, Emans R et al. (2012) Combination of de novo assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in *Schistosoma mansoni*. *Experimental parasitology*, **130**, 470–474.
- Li C, Klco JM, Helton NM et al. (2015) Genetic heterogeneity of induced pluripotent stem cells: results from 24 clones derived from a single C57BL/6 mouse (R Johnson, Ed.). *PLoS ONE*, **10**, e0120585.
- Li F, Mao G, Tong D et al. (2013) The histone mark H3K36me3 regulates Human DNA mismatch repair through its interaction with MutSa. *Cell*, **153**, 590–600.
- Liu T, Rechtsteiner A, Egelhofer TA et al. (2011) Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Research*, **21**, 227–236.
- Lovejoy AO, Stanlis PJ (2011) *The great chain of being: a study of the history of an idea*. Transaction Publishers.
- Low DA, Weyand NJ, Mahan MJ (2001) Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infection and Immunity*, **69**, 7197–7204.
- Lyko F, Foret S, Kucharski R et al. (2010) The honey bee epigenomes: differential methylation of brain DNA in queens and workers (L Keller, Ed.). *PLoS Biology*, **8**, e1000506.
- M** MacIntyre RJ, Clegg MT (2013) *Evolutionary Biology*. Springer Science & Business Media.
- Maher B (2008) Personal genomes: The case of the missing heritability. *Nature*, **456**, 18–21.
- Manolio TA, Collins FS, Cox NJ et al. (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.
- Marchetti M, Jauneau A, Capela D et al. (2014) Shaping bacterial symbiosis With legumes by experimental evolution. *Molecular Plant-Microbe Interactions*, **27**, 956–964.

- M** Marco A, Kozomara A, Hui JHL et al. (2013) Sex-Biased Expression of MicroRNAs in *Schistosoma mansoni* (MK Jones, Ed.). *PLoS Neglected Tropical Diseases*, **7**, e2402.
- Marek M, Kannan S, Hauser A-T et al. (2013) Structural basis for the inhibition of histone deacetylase 8 (HDAC8), a key epigenetic player in the blood fluke *Schistosoma mansoni* (TG Geary, Ed.). *PLoS Pathogens*, **9**, e1003645.
- Martins AV (1957) Non-human vertebrate hosts of *Schistosoma haematobium* and *Schistosoma mansoni*. *Bulletin of The World Health Organization*, **18**, 931–944.
- Massicotte R, Angers B (2012) General-purpose genotype or how epigenetics extend the flexibility of a genotype. *Genetics research international*, **2012**, 1–7.
- Matouk CC, Marsden PA (2008) Epigenetic regulation of vascular endothelial gene expression. *Circulation Research*, **102**, 873–887.
- May J, Meyer CG (2003) Chemoresistance in falciparum malaria. *Trends in Parasitology*, **19**, 432–436.
- Maynard Smith J, Price G (1973) The logic of animal conflict. *Nature*, **246**, 15–18.
- Maynard Smith J (1990) Models of a dual inheritance system. *Journal of Theoretical Biology*, **143**, 41–53.
- Maynard JA, Anthony KRN, Marshall PA, Masiri I (2008) Major bleaching events can lead to increased thermal tolerance in corals. *Marine Biology*, **155**, 173–182.
- Mayr E (1982) *The growth of biological thought: diversity, evolution, and inheritance*. Belknap Press of Harvard University.
- McClintock B (1953) Induction of instability at selected loci in maize. *Genetics*, **38**, 579–599.
- McManus DP, Loukas A (2008) Current status of vaccines for schistosomiasis. *Clinical Microbiology Reviews*, **21**, 225–242.
- Meaburn KJ, Misteli T (2007) Cell biology: chromosome territories. *Nature*, **445**, 379–781.
- Mehta IS, Kulashreshtha M, Chakraborty S, Kolthur-Seetharam U, Rao BJ (2013) Chromosome territories reposition during DNA damage-repair response. *Genome Biology*, **14**, R135.
- Mendel JG (1866) *Versuche über Pflanzenhybriden Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, 1865 Abhandlungen: 3–47*.
- Miller GM, Watson S-A, Donelson JM, McCormick MI, Munday PL (2012) Parental environment mediates impacts of increased carbon dioxide on a coral reef fish. *Nature Climate Change*, **2**, 858–861.
- Mitta G, Adema CM, Gourbal B, Loker ES, Theron A (2012) Compatibility polymorphism in snail/schistosome interactions: From field to theory to molecular mechanisms. *Developmental & Comparative Immunology*, **37**, 1–8.
- Moné Y, Gourbal B, Duval D et al. (2010) A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. *PLoS Neglected Tropical Diseases*, **4**, e813.
- Montanini B, Chen P-Y, Morselli M et al. (2014) Non-exhaustive DNA methylation-mediated transposon silencing in the black truffle genome, a complex fungal genome with massive repeat element content. **15**, 1–16.

- M** Morgan DK, Whitelaw E (2008) The case for transgenerational epigenetic inheritance in humans. *Mammalian genome : official journal of the International Mammalian Genome Society*, **19**, 394–397.
- Morgan JAT, DeJong RJ, Snyder SD, Mkoji GM, Loker ES (2001) *Schistosoma mansoni* and *Biomphalaria*: past history and future trends. *Parasitology*, **123**, 211–228.
- Morgan TH (1915) *The mechanism of Mendelian heredity*. Holt.
- Morgan TH (1917) The Theory of the Gene. *The American Naturalist*, **51**, 513–544.
- N** Nashun B, Yukawa M, Liu H, Akiyama T, Aoki F (2010) Changes in the nuclear deposition of histone H2A variants during pre-implantation development in mice. *Development*, **137**, 3785–3794.
- Nätt D (2008) Stress and the offspring: adaptive transgenerational effects of unpredictability on behaviour and gene expression in chickens (*Gallus gallus*). Liköping University.
- Nilsson EE, Skinner MK (2015) Environmentally induced epigenetic transgenerational inheritance of disease susceptibility. *Translational Research*, **165**, 12–17.
- Nilsson E, Larsen G, Manikkam M *et al.* (2012) Environmentally induced epigenetic transgenerational inheritance of ovarian disease (T Shioda, Ed.). *PLoS ONE*, **7**, e36129.
- Norouzitalab P, Baruah K, Vandegehuchte M *et al.* (2014) Environmental heat stress induces epigenetic transgenerational inheritance of robustness in parthenogenetic *Artemia* model. *The FASEB Journal*, **28**, 3552–3563.
- O** Oliveira G, Rodrigues NB, Romanha AJ, Bahia D (2004) Genome and genomics of schistosomes. *Canadian Journal of Zoology*, **82**, 375–390.
- P** Palumbi SR, Barshis DJ, Traylor-Knowles N, Bay RA (2014) Mechanisms of reef coral resistance to future climate change. *Science (New York, N.Y.)*, **344**, 895–898.
- Pan G, Tian S, Nie J *et al.* (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell*, **1**, 299–312.
- Paracer S, Amhadjian V (2000) *Symbiosis : an introduction to biological associations*. Oxford University Press.
- Parker M, Allen T (2011) Does mass drug administration for the integrated treatment of neglected tropical diseases really work? Assessing evidence for the control of schistosomiasis and soil-transmitted helminths in Uganda. *Health Research Policy and Systems*, **9**, 3.
- Pastor WA, Aravind L, Rao A (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature Publishing Group*, **14**, 341–356.
- Pál C, Miklós I (1999) Epigenetic inheritance, genetic assimilation and speciation. *Journal of Theoretical Biology*, **200**, 19–37.
- Pembrey M, Bygren L, Golding J (2013) The nature of human transgenerational responses. In: *Epigenetics and Human Health* (eds Jirtle RL, Tyson FL), pp. 257–271. Springer Berlin Heidelberg.

- P Peng H, Zhang J (2009) Plant genomic DNA methylation in response to stresses: Potential applications and challenges in plant breeding. *Progress in Natural Science*, **19**, 1037–1045.
- Perrin C, Lepesant JM, Roger E et al. (2013) *Schistosoma mansoni* mucin gene (SmPoMuc) expression: epigenetic control to shape adaptation to a new host (M Knight, Ed.). *PLoS Pathogens*, **9**, e1003571.
- Peserico A, Simone C (2011) Physical and functional HAT/HDAC interplay regulates protein acetylation balance. *Journal of Biomedicine and Biotechnology*, **2011**, 1–10.
- Pfennig DW, Wund MA, Snell-Rood EC et al. (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*, **25**, 459–467.
- Pica-Mattoccia L, Archer S, Cioli D (1992) Hycanthone resistance in schistosomes correlates with the lack of an enzymatic activity which produces the covalent binding of hycanthone to parasite macromolecules. *Molecular and Biochemical Parasitology*, **55**, 167–175.
- Pica-Mattoccia L, Cioli D, Archer S (1988) Binding of tritiated hycanthone and hycanthone N-methylcarbamate to macromolecules of drug-sensitive and drug-resistant schistosomes. *Molecular and Biochemical Parasitology*, **31**, 87–96.
- Pica-Mattoccia L, Dias LC de S, Cioli D (1991) Genetic complementation analysis of two independently isolated hycanthone-resistant strains of *Schistosoma mansoni*. *Memorias do Instituto Oswaldo Cruz*, **87 Suppl 4**, 211–214.
- Picart Picolo A (2015) Étude de l'augmentation de la thermotolérance chez le corail tropical *Pocillopora damicornis*. Université de Perpignan Via Domitia.
- Powles SB, Yu Q (2010) Evolution in action: plants resistant to herbicides. *Annual Review of Plant Biology*, **61**, 317–347.
- Price GR (1970) Selection and covariance. *Nature*, **227**, 520–521.
- Price GR (1972) Fisher's 'fundamental theorem' made clear. *Annals of human genetics*, **36**, 129–140.
- Prokopuk L, Western PS, Stringer JM (2015) Transgenerational epigenetic inheritance: adaptation through the germline epigenome? *Epigenomics*, epi.15.36.
- Protasio AV, Tsai IJ, Babbage A et al. (2012) A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni* (KF Hoffmann, Ed.). *PLoS Neglected Tropical Diseases*, **6**, e1455.
- Putnam HM, Gates RD (2015) Preconditioning in the reef-building coral *Pocillopora damicornis* and the potential for trans-generational acclimatization in coral larvae under future climate change conditions. *Journal of Experimental Biology*, **218**, 2365–2372.
- R Raddatz G, Guzzardo PM, Olova N et al. (2013) Dnmt2-dependent methylomes lack defined DNA methylation patterns. *PNAS*, **110**, 8627–8631.
- Raj S, Bräutigam K, Hamanishi ET (2011) Clone history shapes *Populus* drought responses. *Proceedings of the National Academy of Sciences*, **108**, 12521–12526.
- Rankin D (2013) *Plato and the Individual*. Taylor & Francis.

- R Rasmann S, De Vos M, Casteel CL *et al.* (2012) Herbivory in the previous generation primes plants for enhanced insect resistance. *Plant Physiology*, **158**, 854–863.
- Ratel D, Ravanat J-L, Berger F, Wion D (2006) N6-methyladenine: the other methylated base of DNA. *BioEssays*, **28**, 309–315.
- Reece JB (2011) *Campbell Biology*. Benjamin Cummings / Pearson.
- Remigi P, Capela D, Clerissi C *et al.* (2014) Transient hypermutagenesis accelerates the evolution of legume endosymbionts following horizontal gene transfer (JL Dangl, Ed.). *PLoS Biology*, **12**, e1001942.
- Richards CL, Schrey AW, Pigliucci M (2012) Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation (M Vellend, Ed.). *Ecology Letters*, **15**, 1016–1025.
- Riggs AD (1975) X inactivation, differentiation and DNA methylation. *Cytogenetics and cell genetics*, **14**, 9–25.
- Rizzo PJ (2003) Those amazing dinoflagellate chromosomes. *Cell Research*, **13**, 215–217.
- Roberti A, Sala DL, Cinti C (2006) Multiple genetic and epigenetic interacting mechanisms contribute to clonally selection of drug-resistant tumors: Current views and new therapeutic prospective. *Journal of Cellular Physiology*, **207**, 571–581.
- Roberts DA (1983) Acquired resistance to tobacco mosaic virus transmitted to the progeny of hypersensitive tobacco. *Virology*, **124**, 161–163.
- Robertson KD (2005) DNA methylation and human disease. *Nature Reviews Genetics*, **6**, 597–610.
- Rodríguez-Paredes M, Esteller M (2011) Cancer epigenetics reaches mainstream oncology. *Nature Medicine*, **17**, 328–229.
- Roger E, Mitta G, Moné Y *et al.* (2008) Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: New candidates identified by a global comparative proteomics approach. *Molecular and Biochemical Parasitology*, **157**, 205–216.
- Roger J (1997) *The life sciences in eighteenth-century French thought*. Stanford University Press.
- Roquis D, Lepesant JM, Villafan E *et al.* (2013) Exposure to hycanthone alters chromatin structure around specific gene functions and specific repeats in *Schistosoma mansoni*. *Frontiers in genetics*, **5**, 207–207.
- Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I (2007) The role of microorganisms in coral health, disease and evolution. *Nature Reviews Microbiology*, **5**, 355–362.
- Rosenberg E, Sharon G, Zilber-Rosenberg I (2009) The hologenome theory of evolution contains Lamarckian aspects within a Darwinian framework. *Environmental Microbiology*, **11**, 2959–2962.
- Ruden DM (2015) Epigenetics as an answer to Darwin’s “special difficulty,” Part 2: natural selection of metastable epialleles in honeybee castes. *Frontiers in Genetics*, **6**, 1–15.

- R Russo VEA, Cove DJ, Edgar L, Jaenisch R, Salamini F (2012) *Development: genetics, epigenetics and environmental regulation*. Springer Berlin Heidelberg.
- Russo VE, Martienssen RA, Riggs AD (1996) *Epigenetic mechanisms of gene regulation*. Cold Spring Harbor Laboratory Press.
- S Sammarco PW (1982) Polyp bail-out: an escape response to environmental stress and a new means of reproduction in corals. *Marine ecology progress series*. Oldendorf, **10**, 57–65.
- Saravanan V, Godhe A (2010) Genetic heterogeneity and physiological variation among seasonally separated clones of *Skeletonema marinoi* (Bacillariophyceae) in the Gullmar Fjord, Sweden. *European Journal of Phycology*, **45**, 177–190.
- Sarkar S (1999) *From the Reaktionsnorm to the adaptive norm: the norm of reaction, 1909–1960*. Biology and Philosophy.
- Schelling von FWJ (1993) *System of transcendental idealism (1800)*. University of Virginia Press.
- Schübeler D (2015) Function and information content of DNA methylation. *Nature*, **517**, 321–326.
- Schwaiger M, Schonauer A, Rendeiro AF et al. (2014) Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Research*, **24**, 639–650.
- Schweikert G, Cseke B, Clouaire T, Bird A, Sanguinetti G (2013) MMDiff: quantitative testing for shape changes in ChIP-Seq data sets. *BMC Genomics*, **14**, 826.
- Sharon G, Segal D, Ringo JM et al. (2010) Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, **107**, 20051–20056.
- Shen L, Shao N-Y, Liu X et al. (2013) diffReps: Detecting Differential Chromatin Modification Sites from ChIP-seq Data with Biological Replicates (R Mantovani, Ed.). *PLoS ONE*, **8**, e65598.
- Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics*, **15**, 272–286.
- Shoguchi E, Shinzato C, Kawashima T et al. (2013) Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology*, **23**, 1399–1408.
- Siddiqui AA, Siddiqui BA, Ganley-Leal L (2014) Schistosomiasis vaccines. *Human Vaccines*, **7**, 1192–1197.
- Simões MC, Lee J, Djikeng A et al. (2011) Identification of *Schistosoma mansoni* microRNAs. *BMC Genomics*, **12**, 47.
- Singer CJ (1931) *A short history of biology: a general introduction to the study of living things*. Clarendon Press.
- Skinner MK, Guerrero-Bosagna C, Haque MM (2015) Environmentally induced epigenetic transgenerational inheritance of sperm epimutations promote genetic mutations. *Epigenetics*, **10**, 762–771.

- S** Skinner MK, Gurerrero-Bosagna C, Haque MM et al. (2014) Epigenetics and the evolution of Darwin's finches. *Genome Biology and Evolution*, **6**, 1972–1989.
- Slatkin M (2009) Epigenetic inheritance and the missing heritability problem. *Genetics*, **182**, 845–850.
- Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, **8**, 272–285.
- Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, **14**, 204–220.
- Smocovitis VB (1992) Unifying biology: the evolutionary synthesis and evolutionary biology. *Journal of the history of biology*, **25**, 1–65.
- Song C-X, Yi C, He C (2012) Mapping recently identified nucleotide variants in the genome and transcriptome. *Nature Biotechnology*, **30**, 1107–1116.
- Song S, Wientjes MG, Gan Y, Au JL (2000) Fibroblast growth factors: an epigenetic mechanism of broad spectrum resistance to anticancer drugs. *PNAS*, **97**, 8658–8663.
- Spalding M, Ravilious C, Green EP (2001) *World atlas of coral reefs*. University of California Press.
- Spemann H, Mangold H (1923) induction of embryonic primordia by implantation of organizers from a different species. *The International Journal of Developmental Biology*, **45**, 13-38.
- Stabell OB, Ogbebo F, Primicerio R (2003) Inducible defences in *Daphnia* depend on latent alarm signals from conspecific prey activated in predators. *Chemical Senses*, **28**, 141-153.
- Stedman E (1950) Cell specificity of histones. *Nature*, **166**, 780–781.
- Sun J, Wang S-W, Li C (2014) ATP synthase: an identified target gene of bantam in paired female *Schistosoma japonicum*. *Parasitology research*, **114**, 593–600.
- Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, **9**, 465–476.
- Szyf M (2015) Nongenetic inheritance and transgenerational epigenetics. *Trends in Molecular Medicine*, **21**, 134–144.
- T** Talbert PB, Henikoff S (2010) Histone variants — ancient wrap artists of the epigenome. *Nature Reviews Molecular Cell Biology*, **11**, 264–275.
- Tanabe H, Habermann FA, Solovei I, Cremer M, Cremer T (2002) Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications. *Mutation research*, **504**, 37–45.
- Taylor JH (1982) *Genesi Ad Litteram*. Paulist Press.
- Tellier A, Moreno-Gámez S, Stephan W (2014) Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. *Evolution*, **68**, 2211–2224.
- Theron A (1986) Chronobiology of schistosome development in the snail host. *Parasitology Today*, **2**, 192-194.

- T** Theron A, Rognon A, Gourbal B, Mitta G (2014) Multi-parasite host susceptibility and multi-host parasite infectivity: a new approach of the *Biomphalaria glabrata*/*Schistosoma mansoni* compatibility polymorphism. *Infection, Genetics and Evolution*, **26**, 80–88.
- Théron A, Pointier J-P (1985) Recherche des facteurs susceptibles d'empêcher la réalisation du cycle de *Schistosoma mansoni* dans les mares de la Grande Terre de Guadeloupe (Antilles françaises). *Annales de parasitologie humaine et comparée*, **60**, 155-164.
- Thrall PH, Hochberg ME, Burdon JJ, Bever JD (2007) Coevolution of symbiotic mutualists and parasites in a community context. *Trends in Ecology & Evolution*, **22**, 120–126.
- Tie F, Banerjee R, Stratton CA *et al.* (2009) CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development*, **136**, 3131–3141.
- Trojer P, Reinberg D (2007) Facultative heterochromatin: is there a distinctive molecular signature? *Molecular Cell*, **28**, 1–13.
- Tsaftaris AS, Polidoros AN (1999) DNA methylation and plant breeding. *Plant breeding reviews*.
- Turner BM (1993) Decoding the nucleosome. *Cell*, **75**, 5–8.
- Turner BM (2002) Cellular memory and the histone code. *Cell*, **111**, 295-291.
- V** Valentim CLL, Cioli D, Chevalier FD *et al.* (2013) Genetic and molecular basis of drug resistance and species-specific drug action in schistosome parasites. *Science (New York, N.Y.)*, **342**, 1385–1389.
- van der Graaf A, Wardenaar R, Neumann DA *et al.* (2015) Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proceedings of the National Academy of Sciences*, **112**, 6676–6681.
- van Speybroeck L, van de Vijver G, de Waele D (2002) From epigenesis to epigenetics: the case of C.H. Waddington. *Annals of the New York Academy of Sciences*, **981**, 61-81.
- Van Valen L (1973) A new evolutionary law. *Evolutionary theory*, **1**, 1–30.
- Vastenhouw NL, Schier AF (2012) Bivalent histone modifications in early embryogenesis. *Current Opinion in Cell Biology*, **24**, 374–386.
- Verhoeven KJF, Jansen JJ, Van Dijk PJ, Biere A (2009) Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytologist*, **185**, 1108–1118.
- Vidal-Dupiol J (2011) Stress environnementaux chez le corail scléactiniaire *Pocillopora damicornis*: du modèle expérimental à l'identification de marqueurs fonctionnels du stress. École pratique des Hautes-Études - EHPE Paris.
- Vidal-Dupiol J, Ladrière O, Meistertzheim AL *et al.* (2011) Physiological responses of the scleractinian coral *Pocillopora damicornis* to bacterial stress from *Vibrio coralliilyticus*. *Journal of Experimental Biology*, **214**, 1533–1545.
- Voldgorn YI, Adilgereeva EP, Nekrasov ED, Lavrov AV (2015) Cultivation and differentiation change nuclear localization of chromosome centromeres in Human mesenchymal Stem cells. *PLoS ONE*, **10**, e0118350.

- W Waddington CH (1939) *An Introduction to Modern Genetics*. The Macmillan Company.
- Waddington CH (1940) *Organisers and Genes*. The University Press.
- Waddington CH (1942a) Canalization of development and the inheritance of acquired characters. *Nature*, **150**, 563-565.
- Waddington CH (1942b) The Epigenotype. *Endeavour*, **1**, 18–20.
- Waddington CH (1953) Genetic assimilation of an acquired character. *Evolution*, **7**, 118-126.
- Waddington CH (1956) Genetic assimilation of the bithorax phenotype. *Evolution*, **10**, 1-13.
- Waddington CH (1957) *The Strategy of the Genes*. Allen and Unwin Inc.
- Wang X, Weigel D, Smith LM (2013) Transposon variants and their effects on gene expression in *Arabidopsis*. *PLoS Genetics*, **9**, e1003255.
- Ward S (1992) Evidence for broadcast spawning as well as brooding in the scleractinian coral *Pocillopora damicornis*. *Marine Biology*, **112**, 641–646.
- Watanabe T, Fukasawa T (1961) Episome-mediated transfer of drug resistance in *enterobacteriaceae* I: transfer of resistance factors by conjugation. *Journal of Bacteriology*, **81**, 669-678.
- Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
- Weaver ICG, Cervoni N, Champagne FA *et al.* (2004) Epigenetic programming by maternal behavior. *Nature Neuroscience*, **7**, 847–854.
- Weber A, Declerck S (1997) Phenotypic plasticity of *Daphnia* life history traits in response to predator kairomones: genetic variability and evolutionary potential. *Hydrobiologia*, **360**, 89–99.
- Webster JP, Gower CM, Norton AJ (2008) Evolutionary concepts in predicting and evaluating the impact of mass chemotherapy schistosomiasis control programmes on parasites and their hosts. *Evolutionary Applications*, **1**, 66–83.
- Weismann A, Poulton EB, Schönland S, Shipley AE (1891) *Essays upon heredity and kindred biological problems*. Clarendon Press.
- Whitman DW, Agrawal AA (2009) What is phenotypic plasticity and why is it important. In: *Phenotypic plasticity of insects*, 1–63. CRC Press.
- Wilkins J (2006) Species, kinds, and evolution. *Reports of the National Center for Science Education*, **26**, 36–45.
- Wilson EB (1893) Amphioxus, and the mosaic theory of development. *Journal of Morphology*, **8**, 579–638.
- Wolpert L, Wolpert EPBL, Tickle C, Arias AM (2015) *Principles of development*. Oxford University Press.
- Wu CT, Morris JR (2001) Genes, genetics, and epigenetics: a correspondence. *Science (New York, N.Y.)*, **293**, 1103–1105.
- X Xiang H, Zhu J, Chen Q *et al.* (2010) Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology*, **28**, 516–520.

- Y Youngson NA, Whitelaw E (2008) Transgenerational epigenetic effects. *Annual Review of Genomics and Human Genetics*, **9**, 233–257.
- Z Zentner GE, Henikoff S (2013) Regulation of nucleosome dynamics by histone modifications. *Nature Publishing Group*, **20**, 259–266.
- Zhang G, Huang H, Di Liu *et al.* (2015) N6-methyladenine DNA modification in *Drosophila*. *Cell*, **161**, 1–14.
- Zhang Y, Reinberg D (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & Development*, **15**, 2343–2360.
- Zhou L, Wang Y, Gui JF (2000) Analysis of genetic heterogeneity among five gynogenetic clones of silver crucian carp, *Carassius auratus gibelio* Bloch, based on detection of RAPD molecular markers. *Cytogenetic and Genome Research*, **88**, 133–139.
- Zhu L (2015) Altered levels of circulating miRNAs are associated *Schistosoma japonicum* infection in mice. *Parasites & Vectors*, **8**, 1–11.
- Zilber-Rosenberg I, Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews*, **32**, 723–735.
- Zirkle C (1941) Natural Selection before the “Origin of Species.” *Proceedings of the American Philosophical Society*, **84**, 71–123.
- Zuckerkandl E, Pauling L (1962) Molecular disease, evolution and genetic heterogeneity. In: *Horizons in Biochemistry: Albert Szent-Györgyi Dedicatory Volume*, 189–225. Michael Kasha and Bernard Pullman.
- Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*, **109**, 1193–1198.

CHAPITRE IX

ANNEXES

IX.1. ANNEXE 1 : PUBLICATION SUPPLÉMENTAIRE 1

IX.1.1. Mise en contexte

Avant l'arrivée des technologies de séquençage massif, la variabilité génétique d'une souche de *S. mansoni* au laboratoire était estimée par l'analyse microsatellite sur quatorze marqueurs neutres. Par exemple, une de nos souches, *SmBre*, était couramment utilisée pour les expériences pour lesquelles une lignée isogénique était désirée car elle ne montrait aucune diversité sur l'ensemble de ces marqueurs. Une fois que le séquençage Illumina nous a été accessible, il a été décidé de mieux caractériser la variabilité génétique de nos souches de parasites. Un séquençage populationnel a été accompli sur une centaine d'individus adultes, pour chacune des souches *SmBre* et *SmGua*, les plus couramment utilisées dans notre laboratoire.

Ma participation à ce projet s'est située au niveau de la mise en place des outils bioinformatiques nécessaires pour analyser les données de séquençage massif. L'intérêt était de rendre accessible des logiciels en lignes de commandes, souvent cryptiques, aux biologistes sans expérience en bioinformatique. Ces logiciels ont été inclus dans l'instance locale de Galaxy que j'ai installée et optimisée. Il s'agit d'une interface graphique web permettant d'effectuer plusieurs dizaines de types d'analyses en quelques clics de souris. Cette interface est depuis maintenue par notre bioinformaticien Cristian Chapparo, et est accessible à l'adresse <http://bioinfo.univ-perp.fr>. J'ai également participé à l'analyse des données et à leur visualisation sur le navigateur de génome GBrowse que j'ai également mis en place, et qui est consultable à l'adresse <http://genome.univ-perp.fr>.

IX.1.2. Publication associée

Les résultats de l'analyse génétique des populations de nos deux souches de *S. mansoni* ont été publiés dans la revue **PLoS Neglected Tropical Disease** en décembre 2013 sous le titre « *Private selective sweeps identified from next-generation pool-sequencing reveal convergent pathways under selection in two inbred Schistosoma mansoni strains* ». Les figures et tableaux supplémentaires sont disponibles à l'adresse : <http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002591>.

Private Selective Sweeps Identified from Next-Generation Pool-Sequencing Reveal Convergent Pathways under Selection in Two Inbred *Schistosoma mansoni* Strains

Julie A. J. Clément^{1,2}, Eve Toulza^{1,2}, Mathieu Gautier³, Hugues Parrinello⁴, David Roquis^{1,2}, Jérôme Boissier^{1,2}, Anne Rognon^{1,2}, Hélène Moné^{1,2}, Gabriel Mouahid^{1,2}, Jérôme Buard⁵, Guillaume Mitta^{1,2}, Christoph Grunau^{1,2*}

1 Univ.Perpignan Via Domitia, Ecologie et Evolution des Interactions, UMR 5244, Perpignan, France, **2** CNRS, Ecologie et Evolution des Interactions, UMR 5244, Perpignan, France, **3** INRA, UMR CBGP (INRA – IRD – Cirad – Montpellier SupAgro), Montpellier-sur-Lez, France, **4** MGX Montpellier GenomiX, Montpellier, France, **5** CNRS, Institut de Génétique Humaine, UPR 1142, Montpellier, France

Abstract

Background: The trematode flatworms of the genus *Schistosoma*, the causative agents of schistosomiasis, are among the most prevalent parasites in humans, affecting more than 200 million people worldwide. In this study, we focused on two well-characterized strains of *S. mansoni*, to explore signatures of selection. Both strains are highly inbred and exhibit differences in life history traits, in particular in their compatibility with the intermediate host *Biomphalaria glabrata*.

Methodology/Principal Findings: We performed high throughput sequencing of DNA from pools of individuals of each strain using Illumina technology and identified single nucleotide polymorphisms (SNP) and copy number variations (CNV). In total, 708,898 SNPs were identified and roughly 2,000 CNVs. The SNPs revealed low nucleotide diversity ($\pi = 2 \times 10^{-4}$) within each strain and a high differentiation level ($F_{st} = 0.73$) between them. Based on a recently developed *in-silico* approach, we further detected 12 and 19 private (*i.e.* specific non-overlapping) selective sweeps among the 121 and 151 sweeps found in total for each strain.

Conclusions/Significance: Functional annotation of transcripts lying in the private selective sweeps revealed specific selection for functions related to parasitic interaction (*e.g.* cell-cell adhesion or redox reactions). Despite high differentiation between strains, we identified evolutionary convergence of genes related to proteolysis, known as a key virulence factor and a potential target of drug and vaccine development. Our data show that pool-sequencing can be used for the detection of selective sweeps in parasite populations and enables one to identify biological functions under selection.

Citation: Clément JAJ, Toulza E, Gautier M, Parrinello H, Roquis D, et al. (2013) Private Selective Sweeps Identified from Next-Generation Pool-Sequencing Reveal Convergent Pathways under Selection in Two Inbred *Schistosoma mansoni* Strains. PLoS Negl Trop Dis 7(12): e2591. doi:10.1371/journal.pntd.0002591

Editor: Aaron R. Jex, University of Melbourne, Australia

Received: June 16, 2013; **Accepted:** October 30, 2013; **Published:** December 12, 2013

Copyright: © 2013 Clément et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work received funding from the French National Agency for Research (ANR), project ANR-2010-BLAN-1720-01 (EPIGEVOL) ([http://www.agence-nationale-recherche.fr/programmes-de-recherche/recherches-exploratoires-et-emergentes/blanc-generalite-et-contacts/blanc-presentation-synthese-du-projet/?tx_lwmsuivibilan_pi2\[CODE\]=ANR-10-BLAN-1720](http://www.agence-nationale-recherche.fr/programmes-de-recherche/recherches-exploratoires-et-emergentes/blanc-generalite-et-contacts/blanc-presentation-synthese-du-projet/?tx_lwmsuivibilan_pi2[CODE]=ANR-10-BLAN-1720)). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: grunau@methdb.net

Introduction

In addition to their obvious importance as threats to physical and economical well-being, parasites constitute an interesting group of organisms in which to investigate adaptation and selection. Parasites closely interact with their hosts and entirely depend on them for reproduction and survival. Thus, any change in a host population, which for example decreases parasite ability to penetrate host tissue, will reciprocally select for a change in the parasite such as mechanisms favouring evasion of host resistance to infection. Such an evolutionary arms race [1,2] has been well studied in *Schistosoma mansoni* during the interaction with its intermediate host snail [3–6]. *S. mansoni* is a parasitic plathyhelminth

infecting humans in Africa, the Arabian Peninsula, and South America. It is responsible for the most severe parasitic disease after malaria in terms of morbidity [7–10] killing 200,000 people (WHO Technical Report Series 912: prevention and control of schistosomiasis and soil transmitted helminthiasis (WHO, Geneva, 2002)). *S. mansoni*'s life cycle is characterized by the passage through two obligatory hosts. Parasite eggs are emitted with the faeces of the definitive human or rodent host, but can also accumulate in the liver and cause the symptoms of the disease. When the eggs in host faeces come into contact with water, free-swimming larvae (miracidia) hatch and actively seek their specific intermediate host snails. After active penetration through the tegument, the parasite develops via a primary (mother) sporocyst,

Author Summary

Adaptation of parasites to their environment is governed by the principle of selection. Favourable mutations are fixed in populations while deleterious mutations are progressively eliminated. Here, we aimed to find signatures of selection in two strains of *Schistosoma mansoni*, the causative agent of intestinal schistosomiasis. The strains differ in specific characters, in particular in their capacity to infect intermediate host snails. The reason for this is unknown and understanding it could help control the spreading of the disease. Finding footprints of adaptation to different snail hosts would lead to the discovery of genes that are particularly important for the interaction. Since a single parasite does not contain sufficient DNA to be sequenced, we pooled several individuals, sequenced them as a whole analysed them. In the regions under selection we found genes that are indeed linked to the parasitic lifestyle. We also discovered that natural selection led to diversification of genes that are related to proteolysis, the process by which the parasite destroys host tissue. The related proteins are considered good targets for drug development and vaccination. Our results suggest that in natural populations many variants of these genes exist and that they evolve rapidly, which might hamper therapeutic approaches.

then daughter sporocysts releasing the cercariae that infect the vertebrate host. Then, sexual differentiation takes place within this definitive host and the mating of male and female worms leads to new egg production.

In natural populations, snail/schistosome combinations present different levels of compatibility (*i.e.* the ability for the parasite to penetrate and develop in the host) [3,4]. Previous comparative approaches between a Brazilian (BRE) and Guadeloupean (GH2) *S. mansoni* strains showed that while the first is compatible with a sympatric *Biomphalaria glabrata* strain from Brazil, the latter is much less compatible [7]. Compatibility levels of these strains (and several others not presented in this study) are stable after several years of maintenance under laboratory conditions (Supplementary figure S1). This particular feature has made it possible to elucidate partially the molecular basis of the compatibility polymorphism at the global proteomic [3,4,7–11] and epigenetic scales [12]. In addition, these strains present significant differences in several life history traits, such as chronobiology in cercarial emission [11] and number of mother sporocysts [7]. They are therefore ideal models to investigate signatures of selection at the whole genome scale in *S. mansoni* and to elucidate the genetic basis of phenotypic variation for this parasite.

Three major classes of polymorphisms are responsible for variations in the genotype: (i) single nucleotide polymorphisms (SNPs), a modification of the nucleotide information at a single position, (ii) insertions and deletions (indels) and (iii) structural polymorphisms such as copy number variation (CNV), resulting from tandem duplications of genome segments. These variations can produce by chance favourable, neutral or deleterious phenotypic variations leading to greater, equal or lower fitness, respectively. With the recent advent of Next Generation Sequencing (NGS), hundreds of complete eukaryotic genomes are now available together with huge amounts of data on gene expression and genomic polymorphisms. Standard methods from population genetics to detect selection [3,4,7,13,14] can theoretically be applied to large-scale genomic data, but improvements have been needed to take into account the SNP ascertainment process [15].

These improved methods have been successfully applied to answer different biological questions dealing with the adaptive process, such

as biological invasions [11,16], gene selection in human populations [7,17], and domestication [18,19]. An obstacle for small organisms is the amount of DNA required for library generation and sequencing. In this context, pool-sequencing provides an alternative. The method consists of DNA extraction from a large number of individuals from a population (pool) followed by massive sequencing. SNP frequencies and distributions are then extracted from the sequencing data. While in principle straightforward, current methods for the detection of selective sweeps from pooled sequence SNPs had to be adapted. We used pool-HMM [20–22] in the current work. In this statistical method, the inference of selective sweeps (*i.e.* the elimination of standing variation in regions linked to a recently fixed beneficial mutation) is based on the allele frequency spectrum (AFS) assessed in a sliding window along each chromosome. Our approach allowed sequencing a sufficient number of individuals at moderate costs and to apply population genetics approaches without affecting population genetics estimators [23].

CNV detection is often performed by microarray-based methods comparing two genomes [1] but recent developments such as CNV-seq [24] have extended the method to NGS data. It takes advantage of the variation in the number of short reads aligned in a sliding window along each chromosome to assess CNV at the whole genome scale. As differences in CNV within the same species have been recognized to be involved in adaptive evolution [25,26], we used NGS data to investigate the proportion of this type of variation in our comparative approach.

In the case of *S. mansoni*, we benefit from the availability of a genome assembly [27], facilitating the use of NGS-based strategy for the detection of regions under selection. In this work, we have characterized SNPs from whole genome pool-sequencing data and described their distribution before applying the new population genomics method of Boitard *et al.* [21,22] for detecting selective sweeps as signatures of past selection in these two populations. We also describe differences in copy number variations (CNV) between the two populations as another signature of selection. We finally investigated functional aspects of all genomic regions corresponding to either private selective sweeps or structural variation and proposed evolutionary and ecological interpretations to these genetic differences found between and within BRE and GH2 strains at the whole genome scale.

Materials and Methods

Ethics statement

We adhered to national ethical standards established in the writ of February 1st, 2013 (NOR : AGRG1238753A) setting the conditions for approval, planning and operation of establishments, breeders and suppliers of animals used for scientific purposes and controls. The Ministère de l'Agriculture et de la Pêche and Ministère de l'Éducation Nationale de la Recherche et de la Technologie provided permit A 66040 to our laboratory for experiments on animals and certificate for animal experimentation (authorization 007083, decree 87–848) for the experimenters. Housing, breeding and animal care followed national ethical standards.

Microsatellite analysis

DNA of twenty individual worms (10 males and 10 females) of each strain was individually extracted and genotyped using 15 microsatellite markers. Methods for DNA extraction and microsatellite amplifications were previously published [28].

Schistosoma mansoni DNA preparation for high throughput sequencing

Two *Schistosoma mansoni* strains, one collected in Brazil (BRE) and the other in Guadeloupe (GH2) were used in this study. Each

strain was maintained in its sympatric intermediate host (the mollusk *Biomphalaria glabrata*) and in the mouse (*Mus musculus*) or the hamster (*Mesocricetus auratus*) as a definitive vertebrate host. Genomic DNA was isolated from a pool of a hundred adult individuals. Parasite tissues were digested with 300 mg/L Protease K (Merck, Darmstadt, Germany) in 20 mM TRIS pH 8; 1 mM EDTA; 100 mM NaCl; 0.5% SDS at 55°C overnight. DNA was extracted by two successive rounds of phenol/chloroform followed by chloroform extraction. Precipitation of DNA was done by adding an equal volume of isopropanol/sodium acetate at room temperature [29] and DNA was collected by centrifugation at 12000 rpm for 30 min. After washing with 1 mL of 70% ethanol and air drying, DNA was dissolved in 200 µL of ultrapure water. Quality control and quantification were performed using a spectrophotometer (BioPhotometer, Eppendorf AG, Hamburg, Germany). Twenty µg of DNA was sent to the Montpellier (France) GenomiX facility for Illumina sequencing. Both pooled DNA samples thus represented a random sample from BRE and GH2 populations. If not otherwise stated, reagents were purchased from Sigma-Aldrich (St. Louis, USA).

Illumina (pool-) sequencing

Illumina library production. The Illumina TruSeq DNA sample preparation kit (FC-121-2001, Illumina Inc., San Diego, USA) was used according to the manufacturer's six-steps protocol: 1) 1 µg of DNA was sonicated for 20 minutes by using a Bioruptor (power set on "High", cycles of 30 seconds "ON" and 30 seconds "OFF"); 2) Extremities of sonicated DNA fragments were repaired; 3) an A-base was added to these extremities; 4) Universal (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTC-CCTACACGACGCTCTTCCGATCT-3') and strain-specific (C: 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-TGACCAATCTCGTATGCCGTCTTCTGCTTG-3'; IC: 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGAT-GTATCTCGTATGCCGTCTTCTGCTTG-3') adapters were then ligated to these extremities; 5) DNA products were migrated on 2% agarose gel and 500 bp/±25 fragments were purified (DNA insert size of 380 bp/±25); 6) 12 cycles of PCR amplification were performed to select only DNA fragments carrying both universal and strand specific adapters; 6) Libraries were then validated and quantified on a DNA1000 chip on a Bioanalyzer (Agilent) to determine size and concentration.

Illumina library clustering and sequencing conditions. The cluster generation process was performed on cBot (Illumina Inc.) by using the Illumina Paired-End DNA sample preparation kit (FC-102-1001, Illumina Inc.). DNA libraries were denatured using 0.1 N NaOH then diluted to 8 pM. Diluted libraries were first hybridized to the flow cell (one library per lane) and then amplified leading to DNA cluster formation. After DNA linearization and free-extremity blocking, the first sequencing primer was hybridized to the free-extremity of DNA fragments. Both *S. mansoni* strains were paired-end sequenced on the HiSeq 2000 (Illumina Inc.), using the SBS (Sequence By Synthesis) technique. Paired-end sequencing consisted of sequencing of the first extremity of DNA fragment, followed by the reversal of DNA molecules before the sequencing of the second extremity. Fluorescence was acquired by a camera and image analysis was performed by using HiSeq Control Software (Illumina Inc.). Base calling was achieved by using the RTA software (Illumina Inc.), which corrected and transformed the optical signal from a nucleotide base. Reads of 100 bp from both sides of the fragments were thus obtained after this step.

In silico SNP identification

If not otherwise stated, software was used with default parameters. The Fastx-toolkit version 0.0.13 ([\[cshl.edu/fastx_toolkit/index.html\]\(http://cshl.edu/fastx_toolkit/index.html\)\) was used for quality control and initial cleaning of the sequencing reads. Adaptors were first removed and the three last bases were trimmed from reads because they showed global poor quality scores \(Phred quality score <24\) in most reads. We then filtered reads by their quality score and retained only reads for which at least 90% of the bases had a minimum Phred quality score of 24 \(corresponding to less than 1 incorrect base call in 100 and more than 99% of base call accuracy\). Paired-end data were then considered as two subsets of single-end data. Reads were aligned onto the reference genome version 5.2 \[27\] with the Bowtie software version 0.12.7 \[30\]. Because 47% of the *S. mansoni* genome consists of repeated sequences \[31\] we did not allow for alignment of reads that matched more than one time to avoid false positives in subsequent SNP calling. We allowed 2 mismatches and used the "best" and "strata" options. SNP calling was done with Freebayes software version 0.9.5 \(<https://github.com/ekg/freebayes>\) with parameters `-min-coverage -min-alternate-fraction -min-alternate-total -pooled -ploidy 20`. We tested a minimum coverage of 10 or 20 reads per SNP position \(`-min-coverage 10 or 20`\) with a minimum of 2 reads supporting the variant allele \(`-min-alternate-total 2`\) and with a minimum frequency of variant of 0.02 \(`-min-alternate-fraction 0.02`\). By retaining only SNP that met these conditions, we prevented mistakes in SNP discovery due to errors during base calling and we allowed the discovery of rare variants \(up to a frequency of 0.02\). Among all SNPs, we identified rare and frequent SNPs by using a 20% threshold as proposed in other studies \[17,32\]. *In silico* verification of SNP calls was performed by choosing at random 20 SNPs per chromosome and per strain for each minimum coverage value 10 and 20, corresponding to a total of 640 SNPs. We visually checked the number of reads, the number of variant allele and the alternative base identity with the Next Generation Sequence Assembly Visualization software Tablet \[33\]. All further analyses were done on the dataset obtained with the option `-min-coverage 20`.](http://hannonlab.</p>
</div>
<div data-bbox=)

In vitro SNP verification

To confirm SNP calling, we PCR amplified (primers in Table S8) and re-sequenced 14 DNA fragments covering a total of 22 SNPs. We used a specific melting temperature for each primer pair (Table S8). Sequencing was performed for each PCR product in forward and reverse directions by the GATC biotech AG Company (Konstanz, Germany) using Sanger technology. Sequences were aligned on the reference sequence with Sequencher software version 4.5 (Gene Codes Corporation, Ann Arbor, USA) and SNPs were visually confirmed.

Structural variants

We used CNV-seq to detect copy number variation [24] using a 10^{-6} p-value threshold. Briefly, this algorithm uses variation in sequence coverage in a sliding window between genomic reads mapped to a reference genome to detect variations in sequence copy numbers. Plots were drawn using the associated cnv package in R. We manually checked protein-coding genes contained in the 20 largest genomic regions and the 20 regions with the highest difference in copy number between the BRE and GH2 strain by using the coordinates of the regions in the *Schistosoma mansoni* local Gbrowse instance of the genome (<http://genome.univ-perp.fr>).

Characterization of genetic variability

We used standard tools available on a local Galaxy instance [34] to extract information from the pool-sequencing data. Samtools version 0.1.18 [33] was used to produce mpileup files with read coverage information and call quality from the alignment BAM

files. Nucleotide diversity was evaluated using the unbiased Tajima's Π and Watterson Theta estimators proposed by Futschik and Schlotterer [35] and implemented in Popoolation [36]. Briefly, along each chromosome, each estimator was computed in 50 kb sliding windows (with an overlap of 25 kb between consecutive windows) with the following additional options –min-count 2 –min-coverage 4 –max-coverage 400 –min-qual 20. Considering smaller (10 kb) or larger (100 kb) window size did not affect the results. Fixation index (F_{st}) was assessed on SNPs by using Popoolation2 [37]. We considered all SNPs for which at least 6 reads supported the minor allele for both population simultaneously (–min-count 6), and with a coverage ranging between 20 and 200 reads (–min-coverage 20 ; –max-coverage 200). Mean fixation index between BRE and GH2 was calculated as the average of all F_{st} values obtained for individual SNPs.

Whole-genome scan for footprints of selection

In the current version of the genome, the Z and the W chromosomes are put into a single linkage group [27] despite their physical separation into two chromosomes. SNPs could therefore reflect differences between Z and W and not between target and reference genome. For this reason we excluded them from further analyses of selection. To identify footprints of selection on the 7 autosomes of *Schistosoma mansoni* from BRE and GH2 pool sequences, we relied on the approach recently proposed by Boitard *et al.* [21] implemented in the pool_hmm program [22]. The following option were used to run the analyses: -C 1000 -k 1e-10 –pred -t θ_{wv} where θ_{wv} corresponds to the average unbiased Watterson theta estimator of nucleotide diversity for the population of interest (see above). This method relies on the study of the allele frequency spectrum (AFS) within sliding windows along the genomic sequences. The AFS is expected to be distorted in regions subjected to selection. The model allows estimating the probability of each SNP to belong to one of the three possible (hidden) states (i) neutral, (ii) intermediate and (iii) selected. In practice, one of the critical parameter of the model (defined with the -k option) is the transition probability q assumed between states. The larger the q , the less evidence is required for transition to selection and the more sweep candidates will be detected. In our analyses, we used $q = 10^{-10}$ but also tested less stringent values ($q = 10^{-9}$). As shown in Table S9 and as expected, this leads only to a slightly higher number of footprints. We identified overlapping and private selective sweeps between the two strains by comparing their genomic position for each chromosome. As some genomic region swept in one strain could overlap several shorter regions swept in the other strain (and *vice versa*), we counted overlapping regions as the exact number of genome fragments really overlapping (*i.e.* counting the number of shorter instead of counting only the larger region covering them). Private selective sweeps of each strain correspond to all genomic regions that strictly did not overlap any sweeps of the other strain. Identification of known transcripts in private selective sweeps for both strains was performed using the genome coordinates of the regions and the *Schistosoma mansoni* local Gbrowse instance of the genome. Confidence index was calculated for each specific selective sweep as the maximum of $-\log(1-q_i)$ over the window, where q_i is the posterior probability of hidden state “Selection” given after simulations.

Functional characterization of selected genes

To characterize the molecular functions of the transcripts contained in selective sweeps, we first performed functional annotations of a recent *S. mansoni* transcriptome (unpublished results) by using the Blast2GO software [38]. After the blast step, we mapped gene ontology (GO) terms (BRE: NodeScore = 10,

alpha = 0.4; GH2: NodeScore = 15, alpha = 0.2). We then scanned the proteins with Interproscan [39] and performed GO-Enzyme code mapping to improve annotations before running the annotation step. We finally merged results from these three annotation methods before making functional analyses. We checked if our lists of transcripts matched with proteins that were identified in an earlier study comparing these two strains at the proteomic level [3] using tblastn on the 5.2 version of the *S. mansoni* genome. Results were verified by visual inspection on a local Gbrowse instance of the genome and transcriptome. We then performed functional analyses with Blast2GO tools for the list of transcripts from selective sweeps but not for the list of protein-coding genes from CNV data because this latter list was not exhaustive (see the “Structural variants” previous section). Enrichment analysis was performed for each strain by using the Fisher's exact test with the P-value filter mode set at the default value of 0.05 (Bonferroni correction is applied). Combined graphs were then drawn for each strain from two kinds of data: i) the total list of transcripts found in selective sweeps and ii) the reduced list of transcripts obtained after enrichment analysis. In both cases, Score alpha and Seq Filter value were set at the default values of 0.6 and 5 respectively. Graphs displaying process or function were built on the node score criterion set at the value of 10 or 15 depending on the graph complexity obtained. We also compared molecular pathways in which selection was found by building KEGG maps for each strain within the Blast2GO application.

Synonymous vs. non-synonymous SNPs

SNPEff was used to scan synonymous and non-synonymous SNPs in exons of the whole genome of our parasite model. We downloaded the GFF3 file for the latest assembly (v5.2, nov. 2011) from the Sanger centre (ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/GFF/Smansoni_gff_21032012.tar.gz) and modified it using custom scripts so that gff data and fasta data were separated. We then built a SnpEff database using snpEff.jar with “build -gff3”. A total of 13,385 genes and 14,395 transcripts were detected. The snpEff.config file was modified accordingly and snpEff run with the following parameters: -c snpEff.config -i vcf -o txt -upDownStreamLen 5000 -no None -stats.

Statistical analyses

If not otherwise mentioned, statistical analyses were done on R version 2.15.1 [40]. To test if the SNP density was correlated to the chromosome or the strain, we constructed linear models and tested for the significance of parameters ($p > 0.05$) by analyses of variance (ANOVAs).

Data availability

Sequencing reads are available at the NCBI sequence read archive under study accession number SRP016500 (alias PRJNA177787). Freebayes output (SNPs) is available in vcf and mpileup format at http://2ei.univ-perp.fr/?page_id=2007 and <http://methdb.univ-perp.fr/downloads/>.

Results

Choice of strains

Since pool-sequencing is not yet an established technique for detection of selection we reasoned that it would be suitable to use strains with presumably low genetic diversity. We had earlier characterized the proteome of two laboratory strains of *S. mansoni* of different geographic origins [41] and we had characterized their life history traits such as host compatibility in detail [42]. A comparison of the epigenomes of both strains had identified

differences in chromatin structure in several loci, among them a metalloprotease of the neutral endopeptidase (NEP) family potentially involved in immuno-modulation of *B. glabrata* [12]. A preliminary assessment of genetic variation in both strains was performed by sequencing 15 microsatellites previously described [28]. All microsatellites markers of the Brazilian (BRE) strain were fixed and homozygous. In the Guadeloupean (GH2) strain only one marker (SMDO11) showed two distinct genotypes. Among the 14 fixed markers in the two strains, 6 are shared by BRE and GH2. We concluded that both strains were sufficiently genetically homogenous to perform massive sequencing of a population and phenotypically sufficiently well characterized to potentially link sequence variations to phenotypic traits.

Pool-sequencing

Sequencing of BRE and GH2 strains produced roughly 500,000,000 raw clusters each (Table S1). About 50% passed quality checking (Table S1) and were subsequently aligned to the unique sequences of the reference genome (strain NMRI). Unsurprisingly, we successfully aligned only 46.95% and 59.29% of high-quality reads for BRE and GH2 strains, respectively. This is consistent with the large proportion (47%) of repetitive sequences in the *S. mansoni* genome [31]. Sequencing data are available under study accession number SRP016500 (alias PRJNA177787) at the NCBI sequence read archive.

SNP description. After read alignment on the reference genome and SNP calling for each strain separately, we identified 672,467 polymorphic sites for GH2 and 464,746 for BRE. About 0.2% of them were multiple alleles (*i.e.* more than one alternative allele to the reference genome). Density of polymorphic sites was thus 1.35/kb for BRE (1 polymorphism every 738 bp on average) and 1.96/kb (1 polymorphism every 510 bp on average) for GH2. For all subsequent analyses, we only focused on polymorphic sites that were localized on the annotated and linkage-group/chromosome-level assembled parts of the genome. For the present work, we define SNP at the species level as a nucleotide that is variable within the *S. mansoni*'s genome (*i.e.* with at least one alternative allele different from the reference genome) in at least one of the studied populations/strains or individuals. A SNP can thus fall into one of these two following categories: (i) Private SNP, that is a variable nucleotide within a unique population and fixed to the reference allele in other populations and (ii) Shared SNP, a nucleotide which is different from the reference in at least two populations. Alternative allele frequency (AAF) of a private SNP can thus be equal to zero when fixed in the population, but can take all values from zero (excluded) to one, if several alleles for this position are present in the population. Based on this definition, we found a total of 708,898 SNPs, among which 65,490 were shared (9.2%), 253,321 were private to BRE and 390,087 private to GH2 (Table 1). SNPs were also classified in two categories (rare and frequent) based on the AAF with a cut-off value of 0.2 as was previously proposed [17]. We also defined a sub-category of fixed alternative allele (AAF equal to one) within the "frequent" category. The majority of shared SNPs (70.9%) were fixed (Table 1) or frequent, while private SNPs belonged mainly to the rare category with an AAF ≤ 0.2 (Table 1). For both shared and private SNPs, the majority of SNPs that were classified as frequent (almost 90% in each case) corresponded to fixed alleles (Table 1).

Between chromosome distribution. SNP number on each chromosome for both strains (bars in Figure S2) was highly correlated to chromosome length ($r = 0.93$, $p < 0.0001$) but SNP density (dots in Figure S1) significantly differed between chromosomes ($F_{7,7} = 15.76$, $p < 0.001$) and between strains ($F_{1,7} = 346.28$, $P < 0.001$). However,

no differences in SNP density repartition between strains were found as SNP densities are highly correlated ($\rho = 0.97$, $p < 0.001$).

Within chromosome distribution. SNPs were not equally distributed within chromosomes as we observed SNP clustering for each chromosome of each strain, *i.e.* SNPs tend to aggregate preferentially in some regions and to be lacking in others (Figure S3). This non-random distribution of SNPs across the genome was previously observed in other models [43,44] and is described here for the first time for a metazoan parasite.

Moreover, the distribution of SNP clusters seemed at least partly different between BRE and GH2 strains. Some large peaks of SNP number were common to both strains (see around 55,640,000 window position on ZW sex chromosomes linkage-group, Figure S3H) while others were specific to BRE (see around 3,960,000 window position on Chromosome 6, Figure S3F) or to GH2 (see around 1,140,000 window position on Chromosome 7, Figure S3G).

Nucleotide diversity. We excluded sex chromosomes from subsequent analysis based on allele frequency because of the impossibility to assign SNPs specifically to Z or W chromosomes due to their assemblage into a single linkage group in the present version of the genome. Using Popoolation [36] on the 568,046 SNPs located on autosomes only, we estimated nucleotide diversity. Diversity levels are low with Tajima's P_i (and Watterson Theta) indexes of 2.10^{-4} ($2.9.10^{-4}$) for GH2 and $1.8.10^{-4}$ ($2.6.10^{-4}$) for BRE (Table S2 for full details of diversity measures). This corresponds to an effective population size of hundreds to thousands individuals, assuming a mutation rate around 10^{-7} – 10^{-8} typically found in other eukaryotic genomes [45,46]. This population size is consistent with our laboratory conditions in which vertebrate hosts are usually infected with 140 (mice) to 400 (hamster) cercariae.

Population differentiation. Based on the mean fixation index value ($F_{st} = 0.73 \pm 0.41$ sd), BRE and GH2 populations were highly differentiated. More precisely, among the 229,044 SNPs meeting criteria used in this analysis (see Material & Methods section), 143,391 (60.6%) are fixed in one of the two populations ($F_{st} = 1$).

Identification of structural variants. We used CNV-seq (R version 2.10.1, package 'cnv' version 0.2–7.) to detect copy number variations between the two populations [24]. Because our sequences do not come from a single individual but a population, we used a stringent p-value of 10^{-6} to decrease background variation in coverage. We identified 1,474 CNV for autosomes and 529 for W sex chromosome linkage group with a calculated sliding window of 1,554 and 1,726 nucleotides, respectively (Table 2, Figure 1, Table S3). This number of CNVs is probably underestimated due to relatively modest average sequencing depth. We also analysed independently unassembled scaffolds larger than 500 kb. As outlined above, reads that matched the reference genome more than one time were removed before SNP calling. Manual check of regions with CNV confirmed that they correspond to genomic regions containing single-copied protein coding genes (Supplementary Table S3). We are thus confident that CNV that have been detected in this work do not correspond to repeated sequences.

Identification of selective sweeps

We used the pool-HMM method [20–22] to detect selective sweeps from pool-sequencing data. The question of selective sweeps on sex chromosomes was not addressed in this study. We detected a total of 121 and 151 selective sweeps across the 7 autosomes for BRE and GH2, respectively (Figure 2). We counted a total of 146 overlapping regions, and identified 12 and 19 private

Table 1. Distribution of polymorphic sites and single-nucleotide polymorphism (SNP) for Brazilian (BRE) and Guadeloupean (GH2) strains of *Schistosoma mansoni* used in this study.

	BRE	GH2
1 - Total number of polymorphic sites	464,746	672,467
1.1 - Polybase substitution (% of total polymorphic sites)	1,044 (0.2%)	1,412 (0.2%)
2 - Total SNPs located on chromosomes (autosomes + ZW)	708,898 (568,046+140,852)	
2.1 - Shared SNPs (% of total SNP)	65,490 (9.2%)	
2.1.1 - Rare (% of total SNP - % of total shared)	9,744 (1.4%–20.5%)	
2.1.2 - Frequent: (% of total SNP - % of total shared)	55,746 (7.8%–79.5%)	
2.1.2.1 - Fixed (% of total SNP - % of total shared)	46,440 (6.6%–70.9%)	
2.2 - Private SNPs (% of total SNP)	253,321 (35.8%)	390,087 (55.0%)
2.2.1 - Rare (% of total SNP - % of total private)	173,336 (24.5%–68.4%)	283,456 (40.0%–72.7%)
2.2.2 - Frequent: (% of total SNP - % of total private)	79,985 (11.3%–31.6%)	106,631 (15.0%–27.3%)
2.2.2.1 - Fixed (% of total SNP - % of total private)	69,620 (9.8%–27.5%)	94,666 (13.4%–24.3%)

AAF = alternative allele frequency. Rare SNP correspond to $0 < \text{AAF} \leq 0.2$, frequent correspond to $\text{AAF} > 0.2$ and fixed correspond to $\text{AAF} = 1$.
doi:10.1371/journal.pntd.0002591.t001

selective sweeps for BRE and GH2, respectively (Figure 2) which were differently distributed along the 7 chromosomes (Table 3). Most of these private selective sweeps spanned 100 kb to 1 Mb which corresponded to larger regions than those identified with this method for the X chromosome of *Drosophila melanogaster* [21].

Functional annotation and analyses

Structural variants. We manually checked the CNV encompassing the twenty largest chromosomal regions for the presence of protein-coding genes (Table S4). The largest duplicated region in linkage-group ZW correspond to an additional 268 kb in BRE (Figure 1, Figure S4) and contains 8 protein-coding genes (nucleolar protein 56, histidine triad nucleotide binding protein 1, transcription factor 7 2, exosome complex component RRP45, spermatogenesis associated protein 6, junctophilin 2 and two genes encoding hypothetical proteins). The largest duplicated region in autosomes corresponds to an additional 87 kb on chromosome 3 in GH2 (Figure 1, Figure S4) and contains 2 protein-coding genes encoding telomerase components. We also checked the CNV encompassing the twenty chromosomal regions with the highest value of log2 ratio (*i.e.* with the largest differences in copy number between BRE and GH2) (Table S5). Twelve corresponded to regions without protein-coding genes, four corresponded to regions with genes encoding hypothetical proteins and three contained known genes. Particularly, we identified the presence of (i) a member of immunophilins, a large group of proteins with peptidyl prolyl-isomerase activity (PPI-ase) that exhibit high specificity in binding to immunosuppressive agents (such as cyclosporine); (ii) a homolog of SPARC (secreted protein acidic and rich in cysteine), a secreted glycoprotein involved in interactions between cells and extracellular matrix [47], both with a higher copy numbers in BRE, and (iii) a M13 family peptidase (that present more copies in GH2), a type II membrane metallo-endopeptidase acting on extracellular substrates [48], belonging to a family of proteases for which another member was identified to be epigenetically different between the two strains [12]. Among the 40 CNV regions checked, one was found in both the Top20 (by size) and in the Top20 (by log2 ratio) (CNVR_270) (Table S4 and S5). It contains the abovementioned M13 family proteases.

Private selective sweeps. To identify which functions and/or molecular pathways could have been adaptively selected and could thus be responsible for differences in life history traits and proteomes, we here focused on private selective sweeps of each strain (under arrows in Figure 2, Table 3) assuming that shared selective sweeps could be due to the common laboratory environment.

Within the 12 BRE and 19 GH2 private selective sweeps, we found a total of 592 and 791 predicted genes respectively (Table S5A and Table S5B). Visual inspection of both lists of transcripts in regions under selection revealed the presence of different peptidases for BRE and GH2. In particular, we identified several Cathepsin D and Cathepsin B-like peptidases (Table S6) in BRE, two proteins involved in parasitic activity [49,50].

In a previous work [41], proteomes of sporocysts of *S. mansoni* BRE and GH2 strains had been compared and 17 differentially expressed proteins and/or protein isoforms were revealed. We checked if genes encoding these proteins were located in the regions under selection that we identified in this work. Nine of these proteins (S.mMucin-like, now SmPoMuc family) are encoded by members of a gene family and were already characterized in more detail [51]. Since duplicated loci such as these were excluded in the short-read alignment process, these genes were excluded from the present analysis. None of the 7 genes coding for the 8 remaining protein isoforms are located in any region of private selective sweeps either.

Comparison of combined graphs from Blast2GO gene ontology analysis highlighted several biological processes (Supplementary figure S5). Proteolysis and transport appeared as common biological processes selected in each of the two strains, despite the fact that they correspond to non-overlapping lists of transcripts found in private selective sweeps. Specific processes were also identified, such as cell-cell adhesion for the BRE strain or redox reactions and protein phosphorylation for the GH2 strain.

Using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database, we looked for molecular pathways containing enzymes appearing in regions under selection for each strain. We identified 26 and 34 enzymes (distributed over 24 and 22 pathways) for BRE and GH2, respectively (Table S7). We found five common pathways (Purine metabolism, Riboflavin metabolism, Butanoate metabolism, Aminobenzoate degradation and Carbon fixation). Two of them (Riboflavin metabolism and

Table 2. CNV characteristics for autosomes and ZW-linkage group of *Schistosoma mansoni*.

	Optimal window size used by CNV-seq (bp)	Percentage of CNV (%)	CNV total nucleotide content	CNV count	Mean size (bp)	Median size (bp)	Min-max size (bp)
Autosomes	1,554	4.6%	9,236,932	1,474	6,267	18,649	3,107–87,025
W linkage group	1,726	6.7%	4,010,361	529	7,581	15,534	3,451–268,393

doi:10.1371/journal.pntd.0002591.t002

Aminobenzoate degradation) were highlighted by a unique common enzyme (Enzyme code 3.1.3.2 corresponding to a phosphatase) found under selection and present in both pathways. The three other common pathways were highlighted by enzymes differing between BRE and GH2. In specific pathways, we noted the presence of the N-Glycan biosynthesis for BRE (Table S7), involved in the glycosylation of proteins. Differential glycosylation is involved in the generation of SmPoMucs variants and in part is responsible for compatibility with the intermediate host [42,51].

Synonymous vs. non-synonymous SNPs

Along autosomes, we identified 10,982 and 14,811 SNPs in exons for BRE and GH2 respectively. Proportions of non-synonymous (NonSyn-) and synonymous (Syn-) SNPs within and outside swept regions were similarly distributed for both strains (Supplementary figure S6). Interestingly, 49% of them were Non-Syn-SNPs and found within selective sweeps. We also found more than three times more Non-SynSNPs than Syn-SNPs in private selective sweeps.

Discussion

In this study we have sequenced the genomes of two *Schistosoma mansoni* populations, BRE and GH2, originating from Brazil and Guadeloupe, respectively. Next generation pool-sequencing was applied to this metazoan parasite to analyse genomic variations (SNP and CNV) and to scan for selective sweeps. By mapping genome reads of these two strains to the *S. mansoni* NMRI strain reference genome [27,52], we discovered hundreds of thousands of SNPs and used these data to make the first SNP map for *S. mansoni* at the whole genome scale. We made these data available for further analyses.

Although both strains have been maintained in the lab for more than thirty years (corresponding to approximately one hundred full life-cycles), life history traits such as compatibility levels with the intermediate host as well as chronobiology were maintained overtime (Supplementary figure S1) [7,11]. Earlier investigations had concluded that populations maintained in the laboratory rapidly decreased in diversity to become monomorphic based on the analyses of nine neutral microsatellite markers [28]. It was also previously demonstrated that the diversity of *S. mansoni* decreased dramatically after the first life cycles of laboratory maintenance, based on the analyses of 15 microsatellite markers [28]. Our global comparative approach of BRE and GH2 strains was thus fully justified by their high level of inbreeding and their well-described compatibility polymorphism at the phenotypic and molecular scales [3,7,12,41,51,53]. The use of pool-sequencing allowed for evaluating the genome-wide diversity (SNP distribution and density, SNP frequencies and copy number variations) for each population. Intra-population diversity was low (confirming the results of the earlier microsatellite study) based on the whole-genome nucleotide diversity indexes Tajima's Pi and Watterson Theta and compared to other studies considering nucleotide diversity at this scale [54,55]. Moreover, several lines of evidence indicate a high divergence between populations: (i) we found a high number (and thus a high density) of single-base polymorphisms for both strains when compared to the NMRI reference genome (1.35 and 1.96 mutated site per kb), which was clearly higher than the values (*i.e.* between 0.312 and 0.857 per kb) found after the re-sequencing of lab strains from *Entamoeba histolytica*, another human parasite [56]; (ii) a large majority of SNPs (90.8%) we identified were private, either to BRE (35.8%) or GH2 (55.0%) strains, indicating specific polymorphic loci; (iii) mean fixation index was clearly very high, certainly as a result of the only few

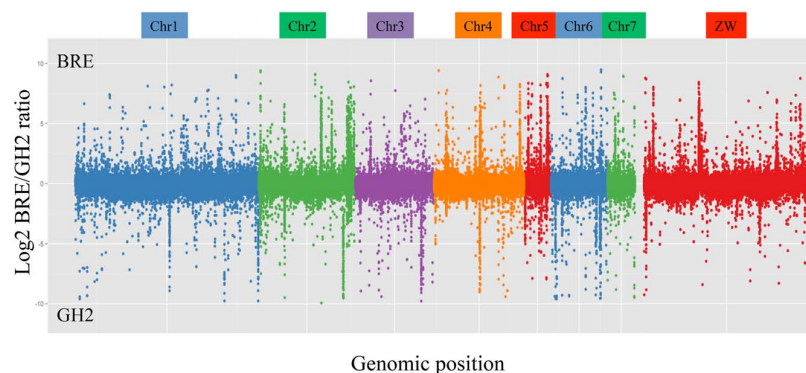


Figure 1. Log 2 ratio plot of copy number variations (BRE/GH2) along the *Schistosoma mansoni* genome. The x-axis represents the genome position in basepairs. Chromosomes are colour coded. The y-axis shows log2 ratio between BRE and GH2. Positive values indicate overrepresentation of a region in BRE, negative values indicate overrepresentation of a region in GH2.
doi:10.1371/journal.pntd.0002591.g001

shared and variable SNPs (Table 1) and (iv) large differences in copy number variations were identified between both strains. All these genome-wide diversity results are fully consistent with the geographic separation of the two strains in the wild since their introduction in the New World from Africa, hampering any genetic exchange, and separate cultivation for decades in the laboratory (Supplementary figure S1). Interestingly, rare SNPs represent the most important class of private SNPs, thus demonstrating that long maintenance in the lab conditions, even if it would largely decrease diversity, could not fully erase segregating polymorphisms.

Using SNP frequencies within each population, we applied a recently published method of population genomics to identify selective sweeps [21,22]. The sweep detection method of Pool-hmm [21,22] uses both the density of segregating sites and the allele frequency pattern among segregating sites to distinguish

sweep regions from neutral regions. In populations with low genetic diversity the length of regions with low genetic diversity in the genome is expected to be higher and therefore mapping resolution lower. Low genetic diversity of our populations will affect the power of the method (false negative sweep detection rate) and the resolution of the sweeps but not its robustness (false positive sweep detection rate). The high number of selective sweeps for both BRE and GH2 strains indicated that a high selective pressure has operated at the whole genome scale. Attributing these signatures of selection to either long evolutionary history or adaptation to laboratory conditions will now require comparison with strains sampled from the field. We therefore focused on private (*i.e.* specific non-overlapping) selective sweeps that may not correspond to adaptation to the common laboratory conditions but to long-term evolutionary processes instead. It also helped to increase resolution because the private selective sweeps

Chromosome 1



Chromosome 2



Chromosome 5



Chromosome 3



Chromosome 6



Chromosome 4



Chromosome 7



Figure 2. Chromosomal representation of selective sweeps (in black) and non-swept regions (grey) over the 7 autosomes of *Schistosoma mansoni* for BRE and GH2. Black arrowheads indicate the non-overlapping specific selective sweeps for both strains.
doi:10.1371/journal.pntd.0002591.g002

Table 3. Selective sweep regions and number of corresponding genes for (A) BRE and (B) GH2 strains.

A	Start	End	Length	Number of genes	CI
Chr1	1,194,541	2,357,392	1,162,852	56	Inf
	61,387,283	62,009,343	622,061	23	11.15
	62,218,658	62,721,763	503,106	19	6.67
	62,877,538	63,098,318	220,781	9	2.71
Chr2	24,145,118	24,315,462	170,345	6	2.83
Chr3	22,823,405	23,467,278	643,874	35	Inf
	24,930,649	25,380,787	450,139	14	Inf
Chr4	20,500,484	21,005,970	505,487	24	8.63
	31,100,629	31,804,241	703,613	36	Inf
Chr5	2,117,276	2,386,209	268,934	17	7.53
	8,814,688	9,017,158	202,471	12	9.05
Chr6	17,508,128	17,649,604	141,477	6	3.94
Chr7	-	-	-	-	-
B	Start	End	Length	Number of genes	CI
Chr1	2,509,484	3,001,529	492,046	43	12.00
	17,514,351	17,728,884	214,534	16	6.36
	44,140,447	44,504,258	363,812	15	4.70
	46,537,352	46,758,483	221,132	13	4.98
	47,463,874	47,701,141	237,268	7	2.83
Chr2	91	101,151	101,061	3	4.07
	832,591	953,172	120,582	8	4.24
	2,627,180	2,753,105	125,926	6	4.74
Chr3	7,872,267	8,528,892	656,626	24	Inf
	9,155,168	9,454,868	299,701	11	11.10
	15,054,008	15,663,195	609,188	52	10.00
	23,728,696	24,321,481	592,786	29	9.51
Chr4	3,816,163	4,152,576	336,414	12	Inf
	16,762,042	17,264,224	502,183	23	Inf
Chr5	7,988,578	8,174,020	185,443	9	3.64
Chr6	18	139,169	139,152	8	9.74
	11,873,433	13,026,934	1,153,502	54	Inf
	15,736,118	15,951,459	215,342	14	3.29
	18,289,818	18,553,369	263,552	21	10.55
Chr7	-	-	-	-	-

Confidence Index (CI) was calculated as the maximum of $-\log(1-q)$ over the window, where q is the posterior probability of hidden state "Selection". (Inf. = infinity). doi:10.1371/journal.pntd.0002591.t003

all constituted smaller regions than other sweeps. We may have here under-estimated the number of regions involved in adaptation and missed some relevant regions that would be identified in further work on field strains. The robustness of the method we used ensures a true and confident discovery of private sweeps, which was confirmed by the high proportion of non-synonymous SNPs that was found within selective sweeps (Supplementary figure S6).

As differences in copy number variation (CNV) within a same species have been recognized to be involved in adaptive evolution [25,26,57], we also investigated such variations as clues of selection between BRE and GH2. We used the recently published CNV-seq method [24] that takes advantage of the high throughput sequencing data and is suitable for pairwise comparisons. The high number (2,003 in total) of CNV found across the whole-genome of

S. mansoni reinforced our hypothesis of high selective pressure and even suggested that these selective pressures were different between both strains and contributed to shape the specific genomic landscape observed for each strain.

Functional annotations and analyses gave more biological significance to these specific signatures of selection. Exploring private selective sweeps, we identified 592 (BRE) and 791 (GH2) transcripts that have been potentially subjected to selection, but some of them could also have been identified because of their genomic proximity with truly selected genes. We therefore regrouped genes in regions under selection and/or with CNV by function using GO terms. Genes that did not form functional groups (outliers) were considered as having little or no significance. Overall, biological processes specific to each strain emerged from functional analysis, among which there is cell-cell adhesion for

BRE and reduction-oxidation reactions, potentially involved in ROS production or ROS scavenging, for GH2. Earlier studies postulated that an evolutionary arms race between snail host and parasite operates preferentially on immune effectors for the Brazilian strain and on immune recognition for the Guadeloupean strain [3,42]. This is consistent with negative selection process acting on these pathways in the two strains. Two pathways involved in N-glycan biosynthesis were also found under selection in BRE. This finding was consistent with previously observed differences between BRE and GH2 in their glycosylation level for SmPoMucs [51]. As SmPoMucs were shown to be involved in compatibility polymorphism, and because the level of glycosylation is directly related to compatibility, we argue that differential selection in the N-glycan biosynthesis pathway may be responsible for this compatibility polymorphism between the two strains. In summary, we find correspondence between observed life-history traits of the different strains and genomic regions under selection. However, we see this rather on the biological function and metabolic pathway level and not on the level of individual genes.

Despite strong genomic divergence between BRE and GH2 strains, functional analyses revealed interesting evolutionary convergence. From non-overlapping lists of transcripts (because only private sweeps were analysed), we highlighted two common biological processes and three common molecular pathways as targets of selection. More particularly, we independently identified numerous proteases in genomic swept regions or in regions showing copy number variations in both strains, indicating an evolutionary convergence to this function. Proteases are indeed key virulence factors of parasites and particularly trematodes, as they are involved in a number of biological processes such as host tissue invasion/migration, nutrition from host substrates (*e.g.* haemoglobin degradation), immune evasion and more generally host-parasite interactions [58,59]. Phylogenomic analysis had shown that protease-encoding genes, including cathepsins for which we found two genes under selection in the BRE strain, were expanded in the *Schistosoma* lineage [60]. Cathepsins are secreted proteases and due to their importance in host-parasite interactions they are considered to be promising targets for the development of novel chemotherapeutic drugs and vaccines against schistosomiasis. Simoes *et al.* [61] previously described SNPs in the cysteine protease Cathepsin B vaccine target gene that are involved in significant conformation changes leading to an alteration in antibody binding to the protein. Variability among *Schistosoma* species at the biochemical level has been described earlier for cathepsin B-like activity [62]. Our results suggest that different selective pressures operated on this gene, and we propose that using this as a target for vaccine development could rapidly become inefficient in natural populations. Cathepsin D is an aspartyl protease originating from successive gene duplication events in the parasite lineage after its diversification from other metazoans and was proposed to be involved in adaptation to the parasitic lifestyle [49]. Here, we also showed that this gene could be under differential selection between strains of the *S. mansoni* species. Further analyses of these two particular genes have to be conducted to investigate the functional basis of this evolutionary signature. From CNV data, we identified an M13 family peptidase, a type II membrane metallo-endopeptidases acting on extracellular substrates [48], belonging to a family of proteases with other member involved in one of the epigenetic differences found between these two strains in a previous study [12]. Characterized members of the family such as neprilysin act on polypeptides smaller than 40 amino-acids. Considering evolutionary relationships among members of the family, it has been proposed that proteases of this family fulfil a broad range of physiological roles [63]. In nematodes, a neprilysin-like protease is involved in locomotion and pharyngeal activity [64]. Combination of our results indicated that the differences in BRE and GH2 life history

traits [3,11] could result from the selection and maintenance of different proteases, leading for example to different compatibility levels with the intermediate host. Altogether, these elements support the hypothesis that the proteolysis function has evolved through different ways within different populations of the *S. mansoni* species. Evolutionary convergence was generally described between species [65] or communities [66] to explain adaptation of different species to a particular life style. Proteolysis could thus be a promising function to further focus on because of its key role in parasitic lifestyle, either to better understand parasite evolution and adaptation to its hosts or to develop new treatment or preventive strategies. Since the function of most genes in the *S. mansoni* genome has not yet been confirmed by functional studies, further investigations are nevertheless needed to objectively demonstrate the above conclusions.

It is important to note that the 47% of repetitive elements in the *S. mansoni* genome were not investigated in this work because we excluded all reads that matched more than one time to the genome to allow SNP and CNV identification. We may have thus under-estimated the number of regions under selection. The recent *de novo* repeat assembly [31] should facilitate such a study on the repetitive part of the genome. The identification of SNPs in segmental duplications could also open new perspectives in *S. mansoni* evolutionary studies. As gene duplication is a mechanism of genomic adaptation to a changing environment [67], it could be largely involved in parasite adaptation as it was suggested earlier with peptidase families [49].

In conclusion, our work provides one of the first examples of a comparative genomic approach based on population sequencing. A genome-wide comparison of single-nucleotide and structural polymorphisms combined with population and functional analyses allowed us here to identify signatures of selection in two *S. mansoni* populations. Even if all the life history traits and proteome complexity have not found their genetic bases through this study, we were able to identify selection acting on specific functions involved in parasitic lifestyle. Notably, our integrative approach highlighted that selection can act on a same function but through different pools of genes, which clearly suggests evolutionary convergence within schistosomes. However, high-throughput genotyping of a larger number of populations (with the pool-sequencing method) is expected to shed more light into evolutionary history and bases of adaptation of *Schistosoma mansoni*. In particular further comparative analyses of field isolates of Old World and New World strains could help to better understand the evolutionary history and diversification of the parasite since its “out-of-Africa” origin [68]. Such future studies would also help us to identify more polymorphisms associated with the intermediate host compatibility and to clarify how the rapid adaptation of the parasite to *Biomphalaria glabrata* has occurred during the colonization of the New World.

Supporting Information

Figure S1 Schematic representation of the history of the two strains used in this study and of the relation between phenotypic and genetic parameters along the chronology of events. On the left (“History”) the known isolation and bottleneck events ① divergence of the two original wild populations from the African origin. ② sampling of individuals leading to the lab-populations for two respective populations ③ maintenance of the two populations in the laboratory for thirty years, in the middle (“Life history traits”) representation of the shifts in life history traits for both strains, Bre in grey, GH2 in black. The during the out of Africa migration is hypothetical, but we know that phenotypic characters remained stable during 30 years in the laboratory. On

the right (“Genetic diversity”), hypothetical shifts in genetic diversity that could have led to the observed F_{ST} and Θ .
(PDF)

Figure S2 Repartition of SNPs number (bars) and density (dots) between chromosomes for the Brazilian (A) and the Guadeloupean (B) strains of *Schistosoma mansoni* used in this study.
(PDF)

Figure S3 Distribution of SNP number in 10 kb-window across each of the seven autosomes (A to G) and across the ZW-linkage group (H) of *Schistosoma mansoni*.
(PDF)

Figure S4 Log 2 ratio plot of copy number variations for the largest CNV genomic region.
(PDF)

Figure S5 Distribution of cellular processes found under selection for the Brazilian (A) and Guadeloupean (B) strains based on their Gene Ontology term at the level 2.
(PDF)

Figure S6 Non-synonymous (NonSyn-) and synonymous (Syn-) SNPs distribution within and outside selective sweeps found in the genome of *Schistosoma mansoni*. Numbers represent the amount of SNPs in each proportion, with the details for specific non-overlapping regions we focused on in this study.
(PDF)

Table S1 Number and proportion of reads conserved after each step of bioinformatic treatment of Brazilian (BRE) and Guadeloupean (GH2) strains of *Schistosoma mansoni*.
(DOCX)

Table S2 Excel file with summary of nucleotide diversity measures of Tajima’s D and Watterson’ Theta for the 7 autosomes of BRE and GH2 strains. See excel file Table_S2_Summary-ResPopulation.xls.
(XLS)

Table S3 Excel file with list of genomic regions with copy number variation between BRE and GH2. See excel file Table_S3_CNV_list.xlsx.
(XLS)

Table S4 Protein-coding genes in the Top 20 CNV regions (by length) in *Schistosoma mansoni* BRE and GH2.
(DOCX)

Table S5 Protein-coding genes in the Top 20 CNV regions (by log2ratio) in *Schistosoma mansoni* strains BRE and GH2.
(DOCX)

Table S6 List of transcripts identified in private selective sweeps for BRE (A) and GH2 (B) strains. See excel file Table_S6_transcripts_in_private_selective_sweeps.xls.
(XLS)

Table S7 Common and specific pathways in which enzymes encoded by genes in genomic regions under selection were found for BRE and GH2 strains.
(DOCX)

Table S8 Primers used for the polymerase chain reaction (PCR) to check *in situ* 22 single nucleotide polymorphisms of both Brazilian and Guadeloupean strains of *Schistosoma mansoni* used in this study.
(DOCX)

Table S9 Number of selective sweeps detected on autosomes of *Schistosoma mansoni* strains BRE and GH2 used in this study for two values of transition probability q .
(DOCX)

Acknowledgments

We would like to thank Julie Lepesant for providing biological material, Jean-François Allienne for technical support during DNA extraction, Alexandre Nore for technical support during *in situ* validation of SNPs, Céline Cosseau for discussion about results and all her precious advice, specifically during *in situ* validation of SNPs, and Simon Boitard for helpful assistance on the implementation of the pool-hmm model. We are grateful to Christopher Bayne for carefully reading the manuscript.

Author Contributions

Conceived and designed the experiments: CG. Performed the experiments: JAJC JBo HP. Analyzed the data: JAJC ET MG JBu CG. Contributed reagents/materials/analysis tools: DR AR HP. Wrote the paper: JAJC ET CG JBu. Initiated and supported the project: GMi. Discussed the results: GMo HM.

References

- Carter NP (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet* 39: S16–S21. doi:10.1038/ng2028.
- Ridley M (2009) *Evolution*. Wiley-Blackwell. 1 pp.
- Roger E, Mita G, Mone Y, Bouchut A, Rognon A, et al. (2008) Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/Schistosoma mansoni model: New candidates identified by a global comparative proteomics approach. *Mol Biochem Parasitol* 157: 205–216. doi:10.1016/j.molbiopara.2007.11.003.
- Theron A, Coustau C (2005) Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? *J Helminthol* 79: 187–191. doi:10.1079/joh2005299.
- Webster JP, Davies CM (2001) Coevolution and compatibility in the snail-schistosoma system. *Parasitology* 123: S41–S56.
- El-Ansary A, Al-Daihan S (2006) Important aspects of *Biomphalaria* snail-schistosoma interactions as targets for antischistosome drug. *Medical Science Monitor* 12: RA282–RA292.
- Theron A, Pages JR, Rognon A (1997) *Schistosoma mansoni*: Distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Exp Parasitol* 85: 1–9. doi:10.1006/expr.1996.4106.
- Chitsulo L, Engels D, Montresor A, Savioli L (2000) The global status of schistosomiasis and its control. *Acta Trop* 77: 41–51. doi:10.1016/s0001-706x(00)00122-4.
- Chitsulo L, Loverde R, Engels D, Barakat R, Colley D, et al. (2004) Schistosomiasis. *Nat Rev Micro* 2: 12–13. doi:10.1038/nrmicro801.
- King CH (2010) Parasites and poverty: The case of schistosomiasis. *Acta Trop* 113: 95–104. doi:10.1016/j.actatropica.2009.11.012.
- Theron A (1980) Mise en évidence de races chonobiologiques de *Schistosoma mansoni*, agent de la bilharziose, à partir de cinétiques d’émission cercarienne. *Comptes Rendus de l’Académie des Sciences, Paris*. 4 pp.
- Lepesant MJ, Grunau C, Cosseau C (2011) Towards an understanding of the epigenetics of schistosomes: a comparative epigenomic study. *Mem Inst Oswaldo Cruz* 106: 823–830.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Suzuki Y (2010) Statistical methods for detecting natural selection from genomic data. *Genes Genet Syst* 85: 359–376. doi:10.1266/ggs.85.359.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, Vol 44. Palo Alto: Annual Reviews, Vol. 39. pp. 197–218. doi:10.1146/annurev.genet.39.073003.112420.
- Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proc Natl Acad Sci USA* 105: 3421–3426. doi:10.1073/pnas.0800107105.
- Akey JM (2002) Interrogating a High-Density SNP Map for Signatures of Natural Selection. *Genome Research* 12: 1805–1814. doi:10.1101/gr.631202.
- Biswas S, Akey JM (2006) Genomic insights into positive selection. *Trends Genet* 22: 437–446. doi:10.1016/j.tig.2006.06.005.
- Gautier M, Flori L, Riebler A, Jaffrezic F, Laloe D, et al. (2009) A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10: 550. doi:10.1186/1471-2164-10-550.
- Boitard S, Schlötterer C, Futschik A (2009) Detecting selective sweeps: a new approach based on hidden markov models. *Genetics* 181: 1567–1578. doi:10.1534/genetics.108.100032.

21. Boitard S, Schlotterer C, Nolte V, Pandey RV, Futschik A (2012) Detecting Selective Sweeps from Pooled Next-Generation Sequencing Samples. *Molecular Biology and Evolution* 29: 2177–2186. doi:10.1093/molbev/mss090.
22. Boitard S, Koller R, Françoise P, Robelin D, Schlotterer C, et al. (2013) Pool-hmm: a Python program for estimating the allele frequency spectrum and detecting selective sweeps from next generation sequencing of pooled samples. *Mol Ecol Resour* 13: 337–340. doi:10.1111/1755-0998.12063.
23. Gautier M, Vitalis R (2013) Inferring population histories using genome-wide allele frequency data. *Molecular Biology and Evolution* 30: 654–668. doi:10.1093/molbev/mss257.
24. Xie C, Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* 10: 80. doi:10.1186/1471-2105-10-80.
25. Lynch M (2002) Genomics - Gene duplication and evolution. *Science* 297: 945–947. doi:10.1126/science.1075472.
26. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155. doi:10.1126/science.290.5494.1151.
27. Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, et al. (2012) A Systematically Improved High Quality Genome and Transcriptome of the Human Blood Fluke *Schistosoma mansoni*. *PLoS Negl Trop Dis* 6: e1455. doi:10.1371/journal.pntd.0001455.
28. Bech N, Beltran S, Portela J, Rognon A, Allienne J-F, et al. (2010) Follow-up of the genetic diversity and snail infectivity of a *Schistosoma mansoni* strain from field to laboratory. *Infection, Genetics and Evolution* 10: 1039–1045. doi:10.1016/j.meegid.2010.06.012.
29. Maniatis T, Sambrook J, Fritsch EF (1982) *Molecular Cloning - A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Press.
30. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25. doi:10.1186/gb-2009-10-3-r25.
31. Lepesant JM, Roquis D, Emans R, Cosseau C, Arancibia N, et al. (2012) Combination of de novo assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in *Schistosoma mansoni*. *Exp Parasitol* 130: 470–474. doi:10.1016/j.exppara.2012.02.010.
32. Marth G, Yeh R, Minton M, Donaldson R, Li Q, et al. (2001) Single-nucleotide polymorphisms in the public domain: how useful are they? *Nat Genet* 27: 371–372. doi:10.1038/86864.
33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. doi:10.1093/bioinformatics/btp352.
34. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, et al. (2005) Galaxy: A platform for interactive large-scale genome analysis. *Genome Research* 15: 1451–1455. doi:10.1101/gr.4086505.
35. Futschik A, Schlotterer C (2010) The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186: 207–218. doi:10.1534/genetics.110.114397.
36. Koller R, Orozco-terwengel P, De Maio N, Pandey RV, Nolte V, et al. (2011) PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE* 6: e15925. doi:10.1371/journal.pone.0015925.
37. Koller R, Pandey RV, Schlotterer C (2011) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27: 3435–3436. doi:10.1093/bioinformatics/btr589.
38. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676. doi:10.1093/bioinformatics/bti610.
39. Zdobnov EM, Apweiler R (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848. doi:10.1093/bioinformatics/17.9.847.
40. R Development Core Team (2005) R: a language and environment for statistical computing.
41. Roger E, Gourbal B, Grunau C, Pierce RJ, Galinier R, et al. (2008) Expression analysis of highly polymorphic mucin proteins (Sm PoMuc) from the parasite *Schistosoma mansoni*. *Mol Biochem Parasitol* 157: 217–227. doi:10.1016/j.molbiopara.2007.11.015.
42. Mitta G, Adema CM, Gourbal B, Loker ES, Theron A (2012) Compatibility polymorphism in snail/schistosome interactions: From field to theory to molecular mechanisms. *Dev Comp Immunol* 37: 1–8. doi:10.1016/j.dci.2011.09.002.
43. Amos W (2010) Even small SNP clusters are non-randomly distributed: is this evidence of mutational non-independence? *Proc R Soc B-Biol Sci* 277: 1443–1449. doi:10.1098/rspb.2009.1757.
44. Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, et al. (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* 24: 381–386. doi:10.1038/74215.
45. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. *Genetics* 148: 1667–1686.
46. Lynch M (2010) Evolution of the mutation rate. *Trends Genet* 26: 345–352. doi:10.1016/j.tig.2010.05.003.
47. Bradshaw AD (2012) Diverse biological functions of the SPARC family of proteins. *Int J Biochem Cell Biol* 44: 480–488. doi:10.1016/j.biocel.2011.12.021.
48. Oefner C, D'Arcy A, Hennig M, Winkler FK, Dale GE (2000) Structure of human neutral endopeptidase (Nephrilysin) complexed with phosphoramidon. *J Mol Biol* 296: 341–349. doi:10.1006/jmbi.1999.3492.
49. Silva LL, Marcet-Houben M, Zerlotini A, Gabaldon T, Oliveira G, et al. (2011) Evolutionary histories of expanded peptidase families in *Schistosoma mansoni*. *Mem Inst Oswaldo Cruz* 106: 864–877.
50. Sajid M, McKerrow JH, Hansell E, Mathieu MA, Lucas KD, et al. (2003) Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol Biochem Parasitol* 131: 65–75. doi:10.1016/s0166-6851(03)00194-4.
51. Roger E, Grunau C, Pierce RJ, Hirai H, Gourbal B, et al. (2008) Controlled Chaos of Polymorphic Mucins in a Metazoan Parasite (*Schistosoma mansoni*) Interacting with Its Invertebrate Host (*Biomphalaria glabrata*). *PLoS Negl Trop Dis* 2: e330. doi:10.1371/journal.pntd.0000330.
52. Berriman M, Haas BJ, Loverde PT, Wilson RA, Dillon GP, et al. (2009) The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460: 352–358. doi:10.1038/nature08160.
53. Cosseau C, Azzi A, Rognon A, Boissier J, Gourbiere S, et al. (2010) Epigenetic and phenotypic variability in populations of *Schistosoma mansoni* - a possible kick-off for adaptive host/parasite evolution. *Oikos* 119: 669–678. doi:10.1111/j.1600-0706.2009.18040.x.
54. Branca A, Paape TD, Zhou P, Briskine R, Farmer AD, et al. (2011) Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proceedings of the National Academy of Sciences* 108: E864–E870. doi:10.1073/pnas.1104032108.
55. Yanagida T, Mohammadzadeh T, Kamhawi S, Nakao M, Sadjjadi SM, et al. (2012) Genetic polymorphisms of *Echinococcus granulosus sensu stricto* in the Middle East. *Parasitol Int* 61: 599–603. doi:10.1016/j.parint.2012.05.014.
56. Weedall GD, Clark CG, Koldjaer P, Kay S, Bruchhaus I, et al. (2012) Genomic diversity of the human intestinal parasite *Entamoeba histolytica*. *Genome Biol* 13: R38. doi:10.1186/gb-2012-13-5-r38.
57. Gemayel R, Vences MD, Legendre M, Verstrepen KJ (2010) Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. *Annu Rev Genet* 44: 445–477. doi:10.1146/annurev-genet-072610-155046.
58. McKerrow JH, Caffrey C, Kelly B, Loke P, Sajid M (2006) Proteases in parasitic diseases. *Annual Review of Pathology-Mechanisms of Disease*. Palo Alto: Annual Reviews, Vol. 1. pp. 497–536. doi:10.1146/annurev.pathol.1.110304.100151.
59. Kasny M, Mikes L, Hampl V, Dvorak J, Caffrey CR, et al. (2009) Peptidases of Trematodes. In: Rollinson D, Hay SI, editors. *Advances in Parasitology*. Vol. 69. pp. 205–. doi:10.1016/s0065-308x(09)69004-7.
60. Silva LL, Marcet-Houben M, Nahum LA, Zerlotini A, Gabaldon T, et al. (2012) The *Schistosoma mansoni* phylome: using evolutionary genomics to gain insight into a parasite's biology. *BMC Genomics* 13: 617. doi:10.1186/1471-2164-13-617.
61. Simoes M, Bahia D, Zerlotini A, Torres K, Artiguenave F, et al. (2007) Single nucleotide polymorphisms identification in expressed genes of *Schistosoma mansoni*. *Mol Biochem Parasitol* 154: 134–140. doi:10.1016/j.molbiopara.2007.04.003.
62. Caffrey CR, Rheinberg CE, Moné H, Jourdan J, Li YL, et al. (1997) *Schistosoma japonicum*, *S. mansoni*, *S. haematobium*, *S. intercalatum*, and *S. rodhaini*: cysteine-class cathepsin activities in the vomitus of adult worms. *Parasitol Res* 83: 37–41.
63. Bland ND, Pinney JW, Thomas JE, Turner AJ, Isaac RE (2008) Bioinformatic analysis of the neprilysin (M13) family of peptidases reveals complex evolutionary and functional relationships. *BMC Evol Biol* 8: 16. doi:10.1186/1471-2148-8-16.
64. Spanier B, Stürzenbaum SR, Holden-Dye LM, Baumeister R (2005) *Caenorhabditis elegans* neprilysin NEP-1: an effector of locomotion and pharyngeal pumping. *J Mol Biol* 352: 429–437. doi:10.1016/j.jmb.2005.06.063.
65. Srithayakumar V, Castillo S, Mainguy J, Kyle CJ (2012) Evidence for evolutionary convergence at MHC in two broadly distributed mesocarnivores. *Immunogenetics* 64: 289–301. doi:10.1007/s00251-011-0588-7.
66. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, et al. (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences* 109: E1878–E1887. doi:10.1073/pnas.1203287109.
67. Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc B-Biol Sci* 279: 5048–5057. doi:10.1098/rspb.2012.1108.
68. Morgan JAT, Dejong RJ, Adego GO, Ansa EDO, Barbosa CS, et al. (2005) Origin and diversification of the human parasite *Schistosoma mansoni*. *Molecular Ecology* 14: 3889–3902. doi:10.1111/j.1365-294X.2005.02709.x.

IX.2. ANNEXE 2 : PUBLICATION SUPPLÉMENTAIRE 2

IX.2.1. Mise en contexte

Depuis le début des années 2000, avec l'avènement du séquençage massif, la communauté scientifique a assisté à une explosion du nombre de génomes d'organismes séquencés. Depuis le début des années 2010, la même chose se produit avec les épigénomes, dont l'importance dans le développement, la réponse à l'environnement et l'évolution adaptative est de plus en plus reconnue. La partie technique associée au décryptage des épigénomes est cependant complexe et fluctue énormément d'un organisme à un autre. Le cas de l'analyse des modifications d'histones par immunoprécipitation de chromatine suivie de séquençage massif (ChIP-Seq) est particulièrement problématique. Les publications de Cosseau *et al.* (2009) et Cosseau & Grunau (2011) ont montré que quelques étapes de validation sont nécessaires avant de pouvoir effectuer du ChIP-Seq sur de nouvelles espèces. Un des points les plus problématiques est la validation des anticorps. Grâce à l'expérience du laboratoire IHPE sur plusieurs invertébrés, nous savons que peu d'anticorps commerciaux ciblant des modifications d'histones donnent satisfaction. Sans ces étapes de validation, il est facile de se laisser induire en erreur par des résultats non-spécifiques.

C'est dans ce contexte que le laboratoire a accueilli Laëtitia Perfus-Barbeoch, de l'UMR 1355 ISA (Université Sophia-Antipolis de Nice), pour l'aider à caractériser l'épigénome d'un nématode parasite des cultures, *Meloidogyne incognita*. Ma contribution à son travail a été de l'aider à mettre au point le ChIP-Seq, en particulier sur les aspects de titrations des anticorps.

IX.2.2. Publication associée

La première description de l'épigénome de *M. incognita* a été publiée dans la revue **Frontiers in Physiology**, en juin 2014, sous le titre « *Elucidating the molecular bases of epigenetic inheritance in non-model invertebrates: the case of the root-knot nematode Meloidogyne incognita* ». Les figures et tableaux supplémentaires sont disponibles à l'adresse : <http://journal.frontiersin.org/article/10.3389/fphys.2014.00211/full>.



Elucidating the molecular bases of epigenetic inheritance in non-model invertebrates: the case of the root-knot nematode *Meloidogyne incognita*

Laetitia Perfus-Barbeoch^{1,2,3*}, Philippe Castagnone-Sereno^{1,2,3}, Michael Reichelt⁴, Sara Fneich^{5,6}, David Roquis^{5,6}, Loris Pratz^{1,2,3}, Céline Cosseau^{5,6}, Christoph Grunau^{5,6} and Pierre Abad^{1,2,3}

¹ INRA, Institut Sophia Agrobiotech, UMR 1355 ISA, Sophia-Antipolis, France

² CNRS, Institut Sophia Agrobiotech, UMR 7254 ISA, Sophia-Antipolis, France

³ Institut Sophia Agrobiotech, Université de Nice Sophia-Antipolis, UMR ISA, Sophia-Antipolis, France

⁴ Abteilung Biochemie, MPI für Chemische Ökologie, Jena, Germany

⁵ Ecologie et Evolution des Interactions, Université de Perpignan Via Domitia, Perpignan, France

⁶ Ecologie et Evolution des Interactions, CNRS, UMR5244, Perpignan, France

Edited by:

Mackenzie Gavery, University of Washington, USA

Reviewed by:

Guilherme Corrêa De Oliveira, Fundação Oswaldo Cruz - FIOCRUZ-Minas, Brazil
Elena Gómez-Díaz, Emory University, USA

*Correspondence:

Laetitia Perfus-Barbeoch, INRA, UMR 1355 ISA, CNRS, UMR 7254 ISA, Institut Sophia Agrobiotech, Université de Nice Sophia-Antipolis, 400 route des Chappes, BP167, Sophia Antipolis cedex, Alpes-Maritimes, France
e-mail: zurletto@sophia.inra.fr

Root-knot nematodes of the genus *Meloidogyne* are biotrophic plant parasites that exhibit different life cycles and reproduction modes, ranging from classical amphimixis to obligatory mitotic parthenogenesis (apomixis), depending on the species. *Meloidogyne incognita*, an apomictic species, exhibits a worldwide distribution and a wide host range affecting more than 3000 plant species. Furthermore, evidences suggest that apomixis does not prevent *M. incognita* from adapting to its environment in contrast to what is expected from mitotic parthenogenesis that should theoretically produce clonal progenies. This raises questions about mechanisms of genome plasticity leading to genetic variation and adaptive evolution in apomictic animals. We reasoned that epigenetic mechanisms might in part be responsible for the generation of phenotypic variants that provide potential for rapid adaptation. We established therefore a pipeline to investigate the principal carriers of epigenetic information, DNA methylation and post-translational histone modifications. Even if *M. incognita* possesses the epigenetic machinery i.e., chromatin modifying enzymes, 5-methyl-cytosine and 5-hydroxy-methyl-cytosine content is absent or very weak. In contrast, we demonstrated that the canonical histone modifications are present and chromatin shows typical nucleosome structure. This work is the first characterization of carriers of epigenetic information in *M. incognita* and constitutes a preamble to further investigate if *M. incognita* development and its adaptation to plant hosts are under epigenetic control. Our pipeline should allow performing similar types of studies in any non-model organism.

Keywords: epigenetics, *Meloidogyne incognita*, chromatin, DNA-methylation, histone modification

INTRODUCTION

Epigenetics is the study of heritable changes in gene expression and function that cannot be explained by changes in DNA sequence. These molecular mechanisms can be influenced by the environment and the individual history as well as being potentially transferable from a generation to the next, with a reversible character (Youngson and Whitelaw, 2008). Organisms are forced to adapt to the heterogeneousness of the environment and the phenotypic plasticity constitutes a major adaptation to compulsory constraints. Since the system of reproduction of an organism conditions its evolutionary dynamics as well as the demography and the spatial genetic structuring of its populations, it is of particular relevance to study epigenetic variation in asexually reproducing organisms. Indeed, in spite of a genetic homogeneity in asexually reproducing organisms, an important phenotypic variability can be observed between individuals of the same clonal lineage found in different environments. The influence of the epigenetic mechanisms could bring an explanation to

this paradox (reviewed in Verhoeven and Preite, 2014). Similarly, epigenetic modifications may provide an accessory source of fast-acting, reversible, and readily available phenotypic variations that can shape host–pathogen interactions (reviewed in Gómez-Díaz et al., 2012). Despite growing evidences on the role of epigenetic phenomena on a wide range of biological phenomena, many questions remain for the involvement of epigenetics in plasticity and adaptation.

Root-knot nematodes (RKN) of the genus *Meloidogyne* are obligatory plant parasites that constitute major agricultural pests worldwide. They infect almost all cultivated plants and establish an intimate interaction with their hosts inducing the re-differentiation of root cells into hypertrophied and multinucleate feeding cells (Caillaud et al., 2008). So far, the most efficient control measure against RKN was the use of chemical substances called nematicides. However, because of their toxic side effects on the environment, nematicides have recently been banned from use. Novel approaches to control them are sorely needed. Among

RKN species, some are strict or facultative amphimictic while others reproduce exclusively by mitotic parthenogenesis (apomixis) (Castagnone-Sereno, 2006). Our model, *Meloidogyne incognita*, reproduces in an asexual way by parthenogenesis without meiosis. Genetically identical individuals develop from *M. incognita* females and form virtually clonal populations. Although these clones share the same genetic heritage, modifications of their phenotype can be observed when they are exposed to unfavorable environments (Castagnone-Sereno et al., 1994). These phenotypes can be completely different such as juveniles molting into either female or male, but also show for instance a switch of virulence leading to infection of resistant hosts. A first step toward a better understanding of plant-nematode interaction was the sequencing of the genome of *M. incognita*. Analysis of this reference genome highlighted characteristics, which could be at the origin of the parasitic success of this species (Abad et al., 2008; Danchin et al., 2010).

Epigenetics changes refer to a set of molecular processes that can affect gene expression such as regulatory processes mediated by remodeling of chromatin structure through methylation of cytosine residues in the DNA and chemical modifications of histone proteins, in particular their acetylation or methylation (Jablónka and Lamb, 2002). On one hand, DNA methylation refers to a chemical modification of genomic DNA by the addition of a methyl (-CH₃) group to specific nucleotide bases. The most common form of DNA methylation is cytosine methylation, occurring predominantly in CpGs in animal genomes (Suzuki and Bird, 2008). However, because of lack of *de novo* methylation machinery, DNA methylation process is unlikely to occur in nematodes, including *C. elegans* but with the exception of *Trichinella spiralis* (Simpson et al., 1986; Wenzel et al., 2011; Gao et al., 2012, 2014). On the other hand, post-translational modifications of histone amino acid side chains (PTMs) are known to mediate epigenetic regulation of gene expression and development in animals (Li et al., 2007; Luger et al., 2012; Turner, 2012). For instance, PTMs, such as acetylation of histone H3 at lysine 9, are typically found in euchromatin and often associated with active transcription, whereas other modifications, such as methylation of lysine 9 or lysine 27 of histone H3, are found in heterochromatin and related to gene repression (Jenuwein and Allis, 2001). One major role for epigenetic variation in evolution would be to promote phenotypic variability, and allow populations to widely explore new environmental conditions. This could lead to a rapid adaptation to environmental changes or colonization of new environments. The interactions between parasites and their host are models of choice to study these mechanisms because the selective pressures are strong and the evolution is fast. *M. incognita* constitutes an ideal model to study these mechanisms, especially as clones can be obtained following the naturally occurring asexual multiplication in host plants. Whereas variation at the genetic level has been studied in *M. incognita*, less is known about the extent and function of epigenetic variation. Previous analyses on multiple *M. incognita* populations have shown that genomic polymorphism could not be directly correlated to observed phenotypes (Semblat et al., 2000).

Many non-model organisms of ecological and economical importance such as *M. incognita* suffer from a lack of even

rudimentary knowledge about their epigenetic information carriers. To address this problem, we describe here a straightforward 2 × 3-step procedure to perform a fundamental analysis of elements of the chromatin marking system. We first performed a whole genome bioinformatics search for chromatin-modifying enzymes such as those involved in DNA methylation, we evaluate the degree of DNA methylation by calculating the CpG observed/expected ratio in *M. incognita* transcriptome and by measuring DNA methylation levels of 5-methylcytosine (5-mC) and 5-hydroxy methylcytosine (5-hmC) in genomic DNA by LC-MS/MS. We then used Western blots to verify the presence of histone and histone modifications and tested critical parameters of a standardized N-ChIP procedure that need to be adapted for each model.

ANALYSIS OF DNA METHYLATION

DNA methylation is a common feature of many but not all genomes. We recommend starting with this feature.

STEP 1: *IN-SILICO* SEARCHES OF CHROMATIN-REMODELING ENZYMES

To understand whether *M. incognita* possesses the capacity to methylate DNA, we first conducted reciprocal BLASTp searches combined to OrthoMCL analysis to identify proteins that might be related to known chromatin-remodeling enzyme orthologs (Table 1). Significant hits were defined as those satisfying the following criteria: *E*-value < 10⁻⁵ and the aligned segments covering at least 30% of the sequence length of the hit. OrthoMCL has been applied to the proteome data set from *M. incognita* and three publicly available nematode genomes (*C. elegans*, *T. spiralis*, and *Pristionchus pacificus*). Animal DNA methyltransferases can be subdivided into three subfamilies, DNMT-1, DNMT-2 and DNMT-3, based on sequence similarity (Kumar et al., 1994; Hendrich and Tweedie, 2003). DNMT-1 has been regarded as a “maintenance” methyltransferase, whereas DNMT-3 functions as “*de novo*” methyltransferase and some evidence suggest that DNMT-2 enzymes are active cytosine-5'-methyltransferases (Kunert et al., 2003). Species exhibiting functional DNA methylation generally encode a complete set of DNMTs in their genomes whereas species lacking DNA methylation, such as *C. elegans*, have lost DNA methylation enzymes from their genomes (Gutierrez and Sommer, 2004; Suzuki and Bird, 2008). BLASTp search of the *M. incognita* whole proteome (Supplementary Material) was performed with DNMT-1, DNMT-2, DNMT-3 protein sequences from nematodes (Table 1): the *Trichinella spiralis* DNMT-1 (accession number EFV58204.1); *P. pacificus* Ppa- DNMT-2 (accession number AY766101.1) and *T. spiralis* DNMT-22 (accession number EFV60295.1); *T. spiralis* DNMT-3 (accession number EFV54759.1). As for the 11 other nematodes in which DNA methylation has been studied, with exception of *T. spiralis* (Gao et al., 2012), *M. incognita* did not possess *de novo* methylation machinery, DNMT-3 nor DNMT-2. However, our data revealed the existence of a potential DNMT-1 orthologous protein annotated in *M. incognita*, as well as methyl-CpG binding domain (MBDs) proteins, which encodes another essential component of the methylation system (Table 1). We found that Minc01117 could be ortholog to the maintenance methyltransferase DNMT-1 from other species. Sites of DNA methylation are occupied

Table 1 | *M. incognita* putative orthologs for methyltransferases (DNMTs) and methyl-CpG binding domain (MBDs) proteins.

Protein	Species	Accession number	<i>M. incognita</i> putative orthologs
DNA METHYLTRANSFERASES (DNMTs)			
DNMT-1	<i>Trichinella spiralis</i>	EFV58204.1	Minc01117
DNMT-2	<i>Trichinella spiralis</i>	EFV60295.1	<i>M. incognita</i> absent from OrthoMCL group
	<i>Pristionchus pacificus</i>	AY766101.1	
DNMT-3	<i>Trichinella spiralis</i>	EFV54759.1	<i>M. incognita</i> absent from OrthoMCL group
METHYL-CpG BINDING DOMAIN 2 (MBDs)			
Cel-MBD-2	<i>Caenorhabditis elegans</i>	C27A12.10	Minc14299 Minc14778
Cbr-MBD-2	<i>Caenorhabditis briggsae</i>	CBP11474	
Ppa-MBD-2	<i>Pristionchus pacificus</i>	AY766102.1	

by various proteins, including MBD proteins, which recruit the enzymatic machinery to establish silent chromatin. Nematodes such as *C. elegans* (Cel-MBD-2, accession number C27A12.10), *C. briggsae* (Cbr-MBD-2, accession number CBP11474) and *P. pacificus* (Ppa-MBD-2, accession number AY766102.1) all contain an mbd-2-like gene (Gutierrez and Sommer, 2004). We found that two predicted *M. incognita* proteins, Minc14299 and Minc14778, could be orthologs to the MBD-2 like protein.

STEP 2: CpG OBSERVED/EXPECTED RATIO (CpGo/e) ANALYSIS

To evaluate the degree of DNA methylation, we calculated the empirical distribution of CpGo/e ratios for *M. incognita* transcriptome. A total of 63,838 expressed sequenced tags (EST) of *M. incognita* that were downloaded from NCBI-dbEST public database (Supplementary Material), filtered by size (≥ 500 bp) leading to 41,649 sequences of which 5000 were randomly chosen for CpG observed/expected ratio (CpGo/e) analysis. Most of the data came from *M. incognita* strain Morelos maintained at INRA Sophia Antipolis and recently published (Abad et al., 2008; Jaouannet et al., 2012). To validate the robustness of CpGo/e ratio method, the same experiment was conducted on *C. elegans* (initially 8722 ESTs), the model nematode lacking DNA methylation process, *D. melanogaster* (38,110 ESTs) and on *A. mellifera* (10,157 ESTs). *D. melanogaster* has been previously described for its lack of DNA methylation at late stages of development (Urieli-Shoval et al., 1982; Lyko et al., 2000) while DNA methylation occurs in *A. mellifera* (Elango et al., 2009). The CpGo/e ratio was used as a proxy to estimate the intragene DNA methylation content as previously described in *Biomphalaria glabrata* (Fneich et al., 2013). In *C. elegans*, the CpGo/e ratio profile exhibited an unskewed Gaussian distribution with an estimated mean value close to 1 (0.93 ± 0.26) as expected for species lacking such process (Figure 1A). A similar result was obtained with

D. melanogaster late stage ESTs (Estimated mean value 0.92 ± 0.17 ; Figure 1B). In contrast, the CpGo/e ratio in *A. mellifera* transcripts can be approximated by a mixture of two Gaussian distributions (Figure 1C). This pattern has also been described in the mollusc *B. glabrata* where transcripts are divided in low and high methylated genes (Fneich et al., 2013). *M. incognita* CpGo/e ratio profile did not perfectly fit *C. elegans* profile (Figure 1D). In *M. incognita*, the CpGo/e ratio profile also exhibited a skewed near-Gaussian distribution but with a slight shift on the left compared to *C. elegans* profile. The estimated mean value for *M. incognita* transcriptome was inferior to 1 (0.77 ± 0.25). However, careful analysis of the curve fitting suggests the presence of two populations of sequences: one with a mean value around 0.9 and another with a mean value around 0.7. These results could indicate the presence of two populations of genes in *M. incognita*: one with no methylation and one with very low methylation.

STEP 3: GLOBAL DNA METHYLATION ANALYSIS BY LC-MS/MS

Genomic DNA used for liquid chromatography-mass spectrometry analysis was prepared from *M. incognita* eggs with the CTAB method (Winneppenninckx et al., 1993) and stored at -80°C until use. The levels of 5-methylcytosine (5-mC) and 5-hydroxy methylcytosine (5-hmC) were determined by liquid chromatography-mass spectrometry LC-MS/MS analysis as their deoxyribonucleosides as described in (Fneich et al., 2013). Neither 5-methyl-cytosine nor 5-hydrox-methyl-cytosine could be detected.

Taken together, our results show that DNA methylation is absent or very low in *M. incognita*.

ANALYSIS OF NUCLEOSOMES, HISTONES, AND HISTONE MODIFICATIONS

In the absence of DNA methylation, PTMs could represent one of the primary epigenetic transcriptional control mechanisms. Such a process has been shown to play a key role in the regulation of gene transcription of parasites, e.g., the apicomplexan family of protozoa and the metazoan *Schistosoma* (Mourão et al., 2012; Perrin et al., 2013). The importance of these modifications in regulating diverse developmental programs in parasites motivated us to study their importance in *M. incognita* in the framework of this study. Experiments were performed on *M. incognita* strain "Morelos" which have been previously used for the whole genome sequencing of this nematode (Abad et al., 2008). Because *M. incognita* is an obligatory plant parasite, it has to be maintained on plants to complete its life cycle. We used tomatoes, *Solanum esculentum* cultivar Saint Pierre, grown at 20°C in a greenhouse as *M. incognita* host for this study. Moreover, samples that we collected are considered to be clonal because of the mitotic parthenogenetic mode of reproduction of *M. incognita* and since the lab population was originated from the progeny of a single female. To grow and produce nematodes, one-month-old tomato plants are inoculated with *M. incognita* second-stage juveniles (J2s). As previously described by Rosso et al., 1999, eggs were collected from 7 week-old infected tomato roots after grinding, sterilizing (0.5% NaOCl) and filtering steps. Extracted eggs were kept either to hatch J2s or to be aliquoted and stored at -80°C for further experiments (e.g., genomic DNA extraction

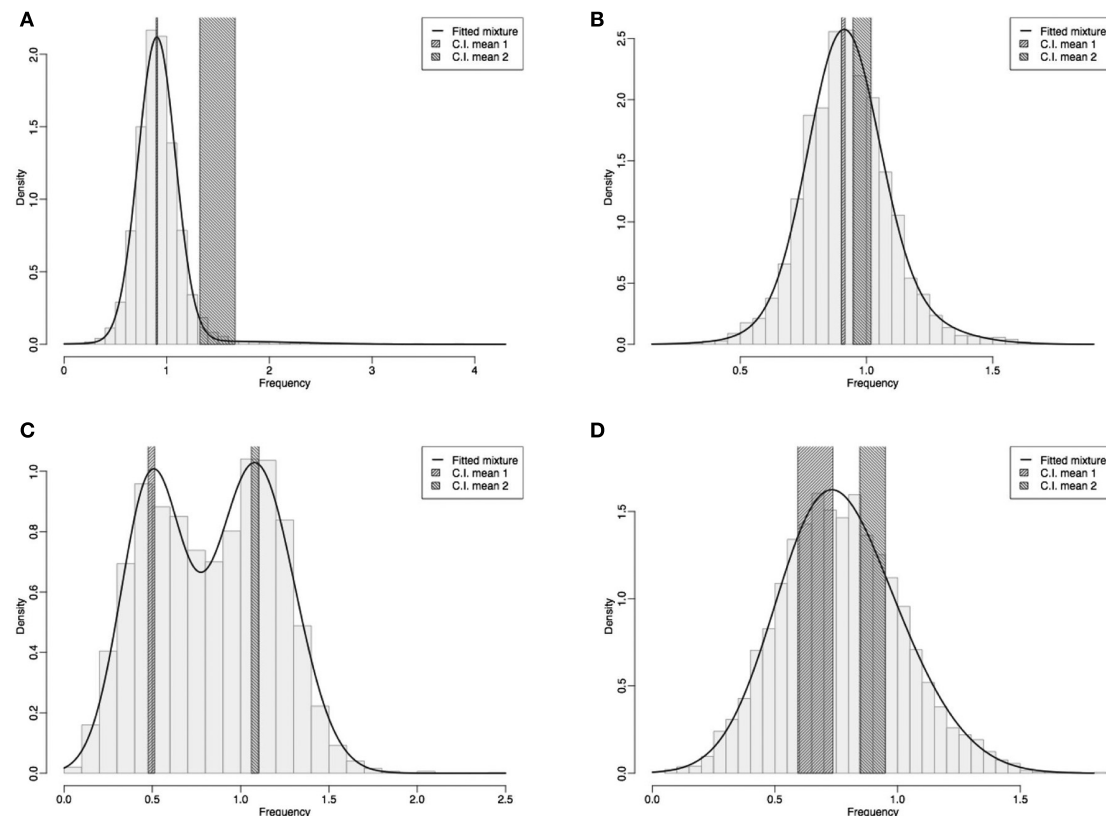


FIGURE 1 | Frequency distribution of CpG observed/expected ratio (CpGo/e) in four different species. (A) *C. elegans* and (B) *D. melanogaster*, and on (C) *A. mellifera* and (D) *M. incognita*. CpGo/e ratio was measured as a proxy to estimate the CpG methylation in transcripts from EST data. X axis:

CpGo/e ratio, Y-axis density (frequency distribution) of ESTs. The figure displays a histogram of Bg GUA CpGo/e ratios with a fitted mixture distribution. The gray shaded bars represent 95% confidence intervals for the two mean values.

and chromatin immuno-precipitation). To produce J2s, extracted eggs were kept on 10 μ m sieve in autoclaved tap water, aerated with an air pump, at room temperature. After 5 days, J2s hatched and were collected by centrifugation (13000 g, 1 min). J2 samples were stored at -80°C for further experiments (e.g., Western blot and chromatin immuno-precipitation).

STEP 1: DETECTION OF CANONICAL HISTONES AND HISTONE MODIFICATIONS BY WESTERN BLOTS

The nucleosome is subject to a dizzying array of posttranslational modifications, which work alone or in combination to constitute a histone code that regulates chromatin structure and function (Jenuwein and Allis, 2001). The four core histones, H2A, H2B, H3, and H4, which are responsible for folding DNA into nucleosomes, have been described as very highly conserved throughout evolution (Marino-Ramírez et al., 2006). This high conservation allowed us to test commercial antibodies that react with mammals but also with a wide range of species. In some species, chromatin is condensed around the centromere and characterized by several methyl groups on lysine amino acid (K) at 9 and 20 position of histone H3 and H4 respectively. These marks are noted H3K9Me2/3 and H4K20Me3 and can be recognized by specific

antibodies. Before immuno-precipitation of *M. incognita* DNA, we tested 12 commercial antibodies, directed against histone H3 and several H3 and H4 modifications, on *M. incognita* J2s with western blots.

Western blots were performed as previously described by Azzi et al. (2009). Briefly, about one thousand *M. incognita* J2s were re-suspended in denaturation buffer, treated by sonication and boiled 5 min at 99°C . Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes (Amersham RPN203D) by the semi-dry method (SEMI-PHOR Bio-Rad). The membrane was blocked overnight at 4°C in blocking buffer and incubated with one of the 12 following commercial antibodies ($1\mu\text{g}/\mu\text{l}$): Anti-histone H3 (H3, abcam ab1791 and Active Motif 39164); Anti-Histone H3 dimethylated at lysine 4 (H3K4Me2, Abcam ab32356) or at lysine 36 (H3K36Me2, Abcam ab9049); Anti-Histone H3 trimethylated at lysine 4 (H3K4Me3, Millipore 04-745 and Abcam ab8580), at lysine 9 (H3K9Me3, Abcam ab8898 and Upstate) or at lysine 27 (H3K27Me3, Diagenode pAB-069-050); Anti-Histone H3 acetylated at lysine 9 (H3K9Ac, Upstate); Anti-Histone H4 trimethylated at lysine 20 (H4K20Me3, abcam ab9053) and Anti-hyperacetylated Histone H4 (H4PentaAc, Upstate). Bands were

revealed by Enhanced Chemical Luminescence (ECL Pierce) and direct exposure to X-ray film (Amersham).

The antibodies should give a unique staining on the membrane (one unique band) to be selected for DNA immuno-precipitation. Three antibodies, Anti-H3K4Me3 (Abcam), Anti-H3K9Me3 (Abcam and Upstate), did not give any signal (**Supplementary Figure 2A**). One antibody, Anti-H4PentaAc gave a non-specific signal with many high molecular weight bands (**Supplementary Figure 2B**). Two antibodies, Anti-H3K27Me3 and Anti-H3K36Me2 gave two distinct bands (**Supplementary Figure 2C**). Six antibodies, Anti-histone H3 (Abcam and Active), Anti-H3K4Me2, Anti-H3K4Me3 (Millipore), Anti-H3K9Ac and Anti-H4K20Me3 gave a unique signal at 17 kDa (**Figure 2**) as expected for histones (Bártová et al., 2005).

We conclude that canonical histones and histone modifications are present and that six commercially available antibodies could be used for immuno-precipitation of *M. incognita* chromatin.

STEP 2: CONTROLLED FRAGMENTATION INTO NUCLEOSOMES

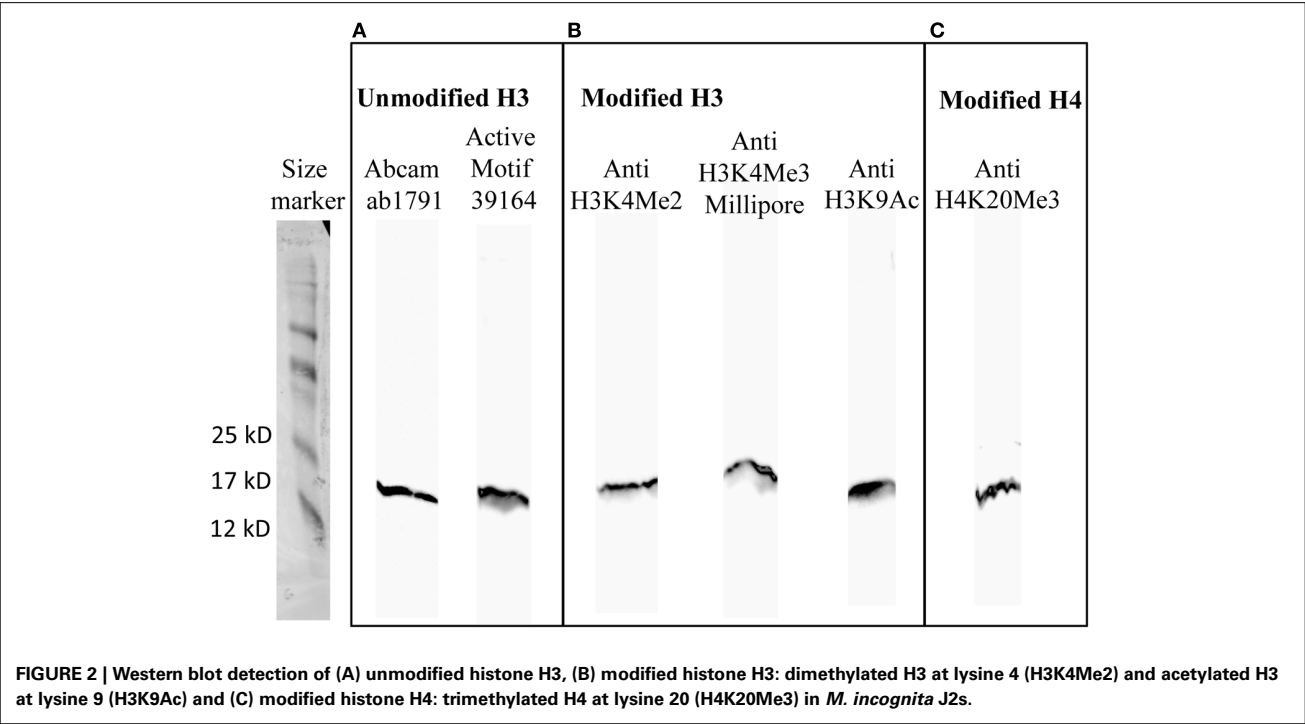
Because little is known about chromatin organization in *M. incognita*, we started by testing partial Micrococcal Nuclease (MNase) digestion of chromatin to reveal its nucleosome structure, which is the fundamental unit of chromatin. Because DNA portions of nucleosome core particles are less accessible for MNase than linking sections, DNA gets digested into fragments of lengths equal to multiplicity of distance between mono-nucleosomes. First, to access nuclei, we tested different grinding times, from 3 to 10 min, for cell lysis in Dounce homogenizer. As a measure for grinding efficiency, we colored nuclei using Hoechst, a

fluorescent stain that binds strongly DNA. Also, because *M. incognita* life-stages are really different in shape and composition, we tested these grinding conditions on both J2s and eggs. Best results were obtained with 10 min total in Dounce homogenizer, distributed as 5 min grinding following by 5 min at rest (**Figure 3A**). Then, we tested different MNase concentrations (from 15 to 30 U; **Figure 3B** and **Supplementary Figure 3**) and incubation times (0, 2, 4, 5, 6, and 7 min). Typical mono-nucleosome structure was observed with fragments about 150 bp and multiplicity of this size (**Figure 3B**). Finally, 5 min incubation with 1 µl MNase (15 U) was chosen as the standard conditions to perform further steps of immuno-precipitation of the chromatin in *M. incognita*.

STEP 3: CHROMATIN IMMUNO-PRECIPITATION (ChIP)

To immunoprecipitate *M. incognita* DNA, we used the ChIP procedure adapted for non-model organisms by Cosseau et al. (2009). Each step of the procedure (**Supplementary Material**) had to be customized for *M. incognita*. Basically, we tested different grinding times for cell lysis and different Micrococcal nuclease (MNase, Affymetrix, USB 70196Y) concentrations and incubation times (see Results). Nuclei were stained with Hoechst 33342 fluorescent stain (1/2000 dilution; Invitrogen H1399) and observed under microscope (40X, Nikon T1-SM) before and after grinding.

Because the antibody should be in excess over the histone of interest for ChIP, we tested protein/chromatin precipitation efficiency in *M. incognita* with increasing antibody volume ranging from 0 to 16 µl (1 µg/µl), for each of 3 tested antibodies (among the 6 antibodies that could be use): Anti-H3 (Abcam), Anti-H3K4Me3 (Millipore), and Anti-H4K20Me3. For each antibody concentration, DNA from the bound and unbound fractions



was extracted with phenol/chloroform so we could quantify the sequences associated with each antibody by qPCR. In other species, histone marks such as H3K4Me3 were enriched around the transcription start sites (TSSs) of genes and were associated with active transcription (Du et al., 2013). For this reason, primers were designed to match genomic region close to TSSs of

M. incognita housekeeping genes, Actin and GAPDH (Minc06773 and Minc12412, respectively; Table 2). Quantitative Real-time PCR (qPCR) analyses were performed using the LightCycler 2.0 system (Roche Applied Science) and LightCycler Faststart DNA Master SYBR Green I kit (Roche Applied Science). qPCR amplification was done in a final volume of 10 μ l composed of 2.5 μ l of immuno-precipitated chromatin, 1.5 μ l H₂O, 0.5 μ l of each primer 10 μ M and 5 μ l of master mix. The following Light-Cycler run protocol was used: denaturation, 95°C, 10 min; amplification and quantification (repeated 40 times), 95°C for 10 s, 62°C for 5 s, 72°C for 16 s; melting curve, 60–95°C with a heating rate of 0.11°C/s and continuous fluorescence measurement, and a cooling step to 40°C. For each reaction, the cycle threshold (Ct) was determined using the “2nd derivative” method of the LightCycler 480 Software release 1.5. Primers were designed for this experiment with the Primer3 web-interface (Table 2). Quality and specificity controls were performed for each qPCR product (Supplementary Figure 1): amplification of a unique 150 bp band was verified by electrophoresis separation through a 2% agarose gel; amplicons were sequenced to verify that they matched expected sequence (GATC-biotech; Table 2); Primer efficiency was calculated based on Ct value for increasing genomic DNA concentration (from 0 to 1 ng/ μ l); Melting curve and negative first derivative of the melting-curve were recorded for each qPCR product to validate that only one product was amplified (one peak was observed).

The amount of target DNA recovered in the immune-precipitated fraction was quantified by calculating the percent input recovery (% IR) normalized with the percent input recovery obtained with the housekeeping gene as previously described by Cosseau et al. (2009).

Also, because of previous MNase digestion, the amplicon size has to be no longer than 150 pb to match *M. incognita* mono-nucleosome length. With *M. incognita* DNA, antibodies Anti-H3 and Anti-H4K20Me3 never reached saturation up to 16 μ l (Figure 4). This could mean that these antibodies are not specific to histone H3 and histone H4 trimethylated at lysine 20 but recognize other proteins/marks. So, the more antibody was added, the more proteins interacted which generated a background signal detected thanks to qPCR. On the contrary, the Anti-H3K4Me3 antibody successfully immuno-precipitated *M.*

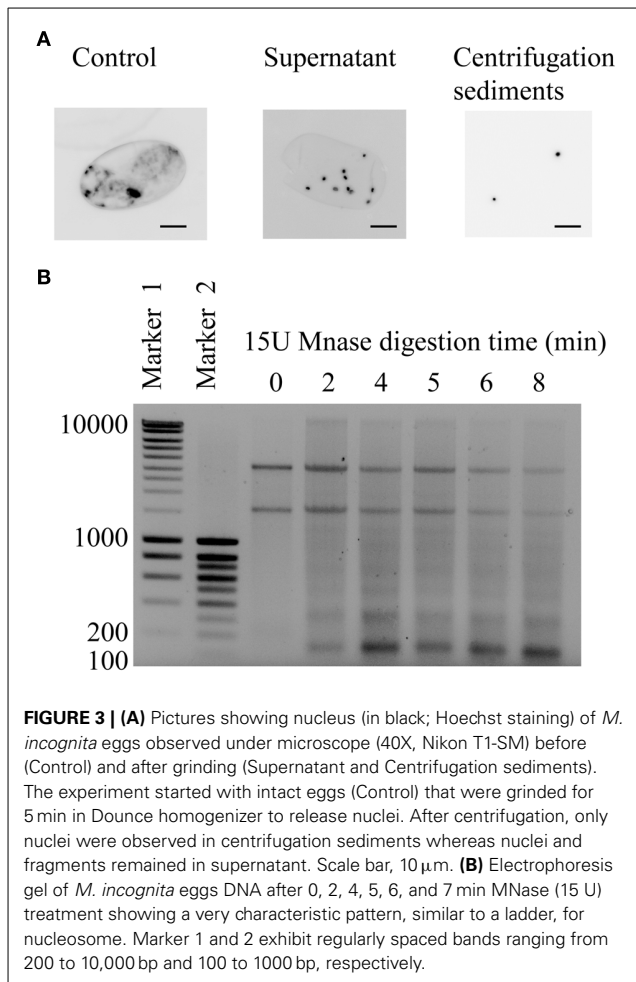


Table 2 | Amplicon sequences and primer sequences designed for qPCR.

Name	Sequence	Length (bp)	Product size (bp)
<i>M. incognita</i> GAPDH amplicon	CGTGCAGCGGTTGAGAAGGATACTGTCCAAGTTGTTGCTGTCAATGACCCGTTTCATTGATCTTGACTATATGTTTGGGGGAGACTTTCTATTAATAACTCAATAACTTTTAGGTCTATATGTTTAACATGATTCCACCCACGGACGC		
Mi_ChIP_GAPDH_F	CGTGCAGCGGTTGAGAAGGA	20	150
Mi_ChIP_GAPDH_R	GCGTCCGTGGGTGGAATCAT	20	
<i>M. incognita</i> Actin amplicon	AAGATGGATGAAGAGGTAGCCGCCCTTGTGTGGATAATGGCAGTGGAATGTGCAAGTTAGTTAAAAATGCCCTTTTCGATTAAAGTTCTGTTGTTTTTAAAGGCTGGATTGCGGGTGATGATGCTCCTCGTGCCGTTTTTCCA		
Mi_ChIP_Actin_F	AAGATGGATGAAGAGGTAGCCGCC	25	150
Mi_ChIP_Actin_R	TGAAAAACGGCAGGAGGACGA	22	

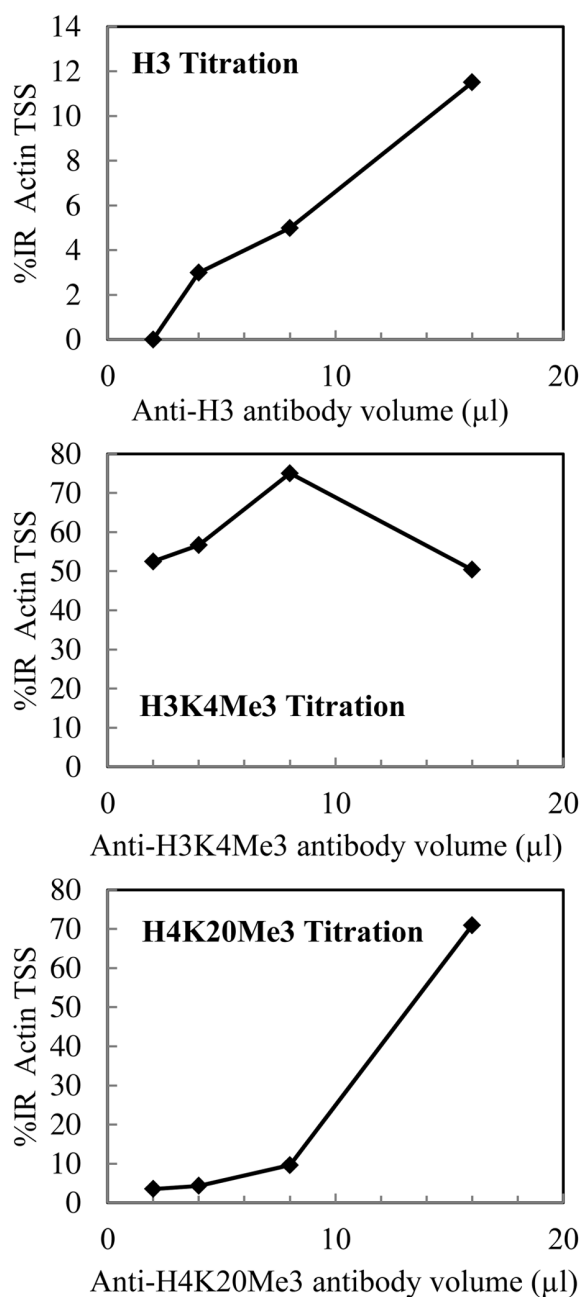


FIGURE 4 | qPCR on *M. incognita* chromatin that had been immuno-precipitated with various volumes (0–16 µl) of Anti-H3, Anti-H3K4Me3, and Anti-H4K20Me3 antibodies (1 µg/µl). The amount of target DNA recovered in the immuno-precipitated fraction was quantified by calculating the percent input recovery (% IR) normalized with the percent input recovery obtained with the housekeeping gene as previously described by Cosseau et al. (2009).

incognita chromatin and reached saturation at 8 µl (Figure 4). Titration saturation for higher antibody concentrations indicates that the antibody is specific to the histone mark (Cosseau et al., 2009). Therefore, this antibody could specifically

immuno-precipitate *M. incognita* and could be used for further analysis.

To summarize, we optimized the ChIP protocol for *M. incognita*, based on the previously published method for non-model organisms (Cosseau et al., 2009). First, we have optimized lysis and extraction of nuclei for *M. incognita* to 10 min grinding in digestion buffer. Then we found that the optimum time for MNase digestion (15U), to obtain fragmented chromatin into mono-nucleosome to penta-nucleosomes, a size range that is optimal for ChIP, was 5 min. Finally, qPCR experiments indicated that N-ChIP procedure should be performed with an excess of 8 µl antibody Anti-H3K4Me3. Now that chromatin immuno-precipitation has been optimized in *M. incognita* it opens new perspectives to further investigate the genome-wide distribution of PTMs by high-throughput sequencing (N-ChIP-Seq).

DISCUSSION

In the past few years, genome projects have demonstrated that DNA methylation is far more widespread than one would expect from the lack of DNA methylation in non-mammalian model organisms, such as the model nematode *C. elegans*. Moreover, functional DNA methylation has been recently reported in the parasitic nematode *T. spiralis* (Gao et al., 2012, 2014). Altogether, this makes us wonder what DNA methylation picture in *M. incognita* could be.

The *de novo* DNA methylation process can produce new methylated sites without pre-existing pattern by affecting cytosine nucleotide (Goll and Bestor, 2005). Depending on CG density, some of these *de novo* sites can then be maintained by DNMT maintenance methyltransferases. CpG under representation is generally seen as a sign of DNA methylation (Bird, 1980). CpG o/e suggests that DNA methylation was present as a subset of genes in *M. incognita* ancestral species, that these sequences were obtained by horizontal gene transfer from a species with DNA methylation or that the EST database was contaminated. In the first case, *M. incognita* might have lost DNA methylation enzymes, which may have been present in the ancestral nematode, from its genome, as what is thought for *C. elegans* (Gutierrez and Sommer, 2004; Suzuki and Bird, 2008). Nevertheless, the observed CpG bias could also be due to other mechanisms (Jabbari and Bernardi, 2004). By LC-MS DNA methylation could not be detected in *M. incognita* that goes in line with the absence of gene coding for *de novo* DNA methylation activity.

The nucleosome is one of the most highly conserved structures known in eukaryotes, which is consistent with the particle's fundamental role in packaging DNA into the nucleus (Turner, 1993, 2005). After prolonged digestion, the MNase degrades the chromatin to its fundamental repeating unit, called the nucleosome. In *M. incognita*, DNA fragment observed micrococcal nuclease digestion of 150 bp defines the nucleosome that is consistent with what is known of the "primary structure" of chromatin with DNA fragment of length around the ~145–150 pb (Kornberg and Thomas, 1974; Luger et al., 1997; Rando, 2011). However, it is becoming clear that chromatin structures are not nearly as uniform and regular as previously assumed: nucleosomes vary in their histone protein components due to the incorporation of variant histones and PTMs (Luger et al., 2012;

Turner, 2012). All PTMs, reversible for the most part, either change nucleosome structure directly by affecting histone proteins and DNA interactions or indirectly by recruiting binding proteins that act on the underlying chromatin structure, as has been proposed in the “histone code” hypothesis (Strahl and Allis, 2000). However, in spite of the high conservation of histone proteins along evolution, their degree of modification varies enormously according to the species, the developmental stage, the tissue and the cellular cycle phase (Strahl and Allis, 2000; Trojer and Reinberg, 2007; Luger et al., 2012; Turner, 2012). Here, we provided evidences for a “canonical nucleosome structure” in *M. incognita* with recognition of both unmodified and modified histone H3 as well as modified histone H4. Both euchromatin and heterochromatin regions contain specific PTMs and binding proteins. For instance, in *C. elegans*, heterochromatin is enriched in trimethylated H3 histones at lysine 56 (H3K56Me3) and lysine 9 (H3K9Me3; Jack et al., 2013). Ultimately, to identify features of transcriptional networks that regulate developmental processes, ChIP-sequencing was completed genome-wide in *C. elegans* (Niu et al., 2011). We expect to build on this foundation, together with this preliminary study, for further evaluation of transcriptional regulatory mechanisms during *M. incognita* development.

In addition, it is to be noted that nematodes could be divided in monocentric species and holocentric species (Subirana and Messeguer, 2013). *C. elegans* as well as *M. incognita* belong to the latter in which the centromeric function spreads out over the whole chromosomes (Triantaphyllou, 1983; Dernburg, 2001; Melters et al., 2012). Moreover, changes in chromosome number and structural rearrangements have been reported in *Meloidogyne* species (Triantaphyllou, 1981; Castagnone-Sereno, 2006; Abad et al., 2008). Being able to study position and histone protein components of *M. incognita* nucleosomes can be used to better understand diverse biological processes including DNA replication, recombination, mutation and repair. Because of the intimate relationship between nucleosome locations and these processes, we anticipate that ChIP-seq experiments in *M. incognita* will shed more light on genome evolution that seems to be crucial for an organism whose mode of reproduction is obligatory mitotic parthenogenesis (apomixis).

One major role for epigenetic variation in evolution would be to promote phenotypic variability, and allow populations to widely explore new environmental conditions. However, the detailed mechanisms thanks to which parasites evade host immunity for instance are poorly understood. *M. incognita* constitutes an ideal model to study these mechanisms, especially as inheritance of the nematode virulence does not segregate in accordance with Mendel’s laws together with asexual multiplication in host plants (Castagnone-Sereno et al., 1994). Because clones can be obtained, genetic variability is reduced. As a result, *M. incognita* provides support for the study of epigenetic heredity distinct from the canonical Mendelian rules. Recently, it became apparent that parasitism success is partly accomplished by epigenetic means in several parasites, including, among others, protozoans *Plasmodium* and trematodes *Schistosoma*. More especially, PTMs plays a key role in the regulation of gene transcription of these parasites (Mourão et al., 2012; Perrin et al., 2013). For

instance, it has been conclusively shown that PTMs, especially acetylations and methylations, and the propagation of heterochromatin away from the telomeres control the transcriptional switching between VAR genes that will trigger antigenic variation and enable immune evasion by the parasite, *Plasmodium falciparum* (Freitas-Junior et al., 2005; Chookajorn et al., 2007; Malmquist et al., 2012; Jiang et al., 2013). In *Schistosoma mansoni*, N-ChIP-seq experiments have been successfully performed to compare chromatin structure at different stages of the parasite life cycle, on different strains and on different sexes (Caby and Pierce, 2009; Cosseau et al., 2009; Lepesant et al., 2011, 2012). These advances in understanding *S. mansoni* epigenetic regulation have lead to the identification of promising new targets for the development of new drug to combat schistosomiasis (Pierce et al., 2012; Lancelot et al., 2013; Marek et al., 2013).

The importance of PTMs in regulating crucial developmental processes in parasites combined with the recent resolution of the genome sequences of *M. incognita* make the study of epigenetic machinery possible and susceptible to break new ground.

CONCLUSION

For the first time in *M. incognita*, this study details DNA methylation and nucleosome structure, carriers of epigenetic information. Genomic DNA of *M. incognita* is not methylated. The general characteristics of nucleosome structure that we observed in the *M. incognita* genome are in accordance to what epigenetic studies have reported from other invertebrate species. Histone modifications, important markers of function and chromatin state, were identified in *M. incognita*. This study opens the way for analyzing epigenetic mechanisms in *M. incognita* at a whole-genome scale to identify new biological processes involved in the generation of phenotypic variation in absence of sexual reproduction in an ecological and economical important model.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: Céline Cosseau, Christoph Grunau, Laetitia Perfus-Barbeoch, Pierre Abad, Philippe Castagnone-Sereno. Performed the experiments: David Roquis, Laetitia Perfus-Barbeoch, Lorix Pratz, Michael Reichelt, Sara Fneich. Analyzed the data: Céline Cosseau, Christoph Grunau, Laetitia Perfus-Barbeoch, Pierre Abad, Philippe Castagnone-Sereno. Wrote the paper: Céline Cosseau, Christoph Grunau, David Roquis, Laetitia Perfus-Barbeoch, Michael Reichelt, Pierre Abad, Philippe Castagnone-Sereno, Sara Fneich.

ACKNOWLEDGMENTS

The authors want to thank Nathalie Marteu (INRA, Sophia Antipolis, France) for growing plant and nematode materials, Etienne G. J. Danchin (INRA, Sophia Antipolis, France) for help with OrthoMCL analysis and Bruno Favery (INRA, Sophia Antipolis, France) for critical reading of the manuscript. This work was supported by the French Government (National Research Agency, ANR) through the “Investments for the Future” LABEX SIGNALIFE: program reference # ANR-11-LABX-0028-0.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fphys.2014.00211/abstract>

Supplementary material : Links to public databases.

Expressed sequenced tags (EST) of *M. incognita* from NCBI-dBEST: <http://www.ncbi.nlm.nih.gov/nucest/?term=meloidogyne+incognita>

Primer3 web-interface: www.bioinformatics.nl/primer3plus/

M. incognita whole proteome: http://www6.inra.fr/meloidogyne_incognita/Genomic-resources/Blast

Supplementary Figure 1 | Quality and specificity controls for qPCR

analysis. (A) Electrophoresis gel of 150 bp genomic DNA amplified with primers specifically designed for ChIP validation. Markers 1 and 2 exhibit regularly spaced bands ranging from 200 to 10,000 bp and 100 to 1000 bp, respectively. **(B)** Primer efficiency was assessed on increasing genomic DNA concentration (from 0 to 1 ng/ul): increase in fluorescence intensity is proportional to the increase in amplicon concentration. **(C)** Melting curves and negative first derivative of the melting-curve: presence of a single homogeneous melt peak confirms specific amplification.

Supplementary Figure 2 | Western blots showing non-specific signals in *M. incognita* J2s. (A)

Anti-H3K4Me3 or Anti-H3K9Me3 antibodies. **(B)** Anti-H4PentaAc antibody. **(C)** Anti-H3K27Me3 or Anti-H3K36Me2 antibodies.

Supplementary Figure 3 | Electrophoresis gel of *M. incognita* eggs DNA

after 0, 2, 4, 5, 6, and 7 min MNase (30U). Marker 1 and 2 exhibit regularly spaced bands ranging from 200 to 10,000 bp and 100 to 1000 bp, respectively.

REFERENCES

- Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., et al. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.* 26, 909–915. doi: 10.1038/nbt.1482
- Azzi, A., Cosseau, C., and Grunau, C. (2009). *Schistosoma mansoni*: developmental arrest of miracidia treated with histone deacetylase inhibitors. *Exp. Parasitol.* 121, 288–291. doi: 10.1016/j.exppara.2008.11.010
- Bártová, E., Pacherník, J., Harnicarová, A., Kovarik, A., Kovariková, M., Hofmanová, J., et al. (2005). Nuclear levels and patterns of histone H3 modification and HP1 proteins after inhibition of histone deacetylases. *J. Cell Sci.* 118, 5035–5046. doi: 10.1242/jcs.02621
- Bird, A. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* 8, 1499–1504. doi: 10.1093/nar/8.7.1499
- Caby, S., and Pierce, R. J. (2009). Quantitative chromatin immunoprecipitation (Q-ChIP) applied to *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 166, 77–80. doi: 10.1016/j.molbiopara.2009.02.014
- Caillaud, M. C., Dubreuil, G., Quentin, M., Perfus-Barbeoch, L., Lecomte, P., de Almeida Engler, J., et al. (2008). Root-knot nematodes manipulate plant cell functions during a compatible interaction. *J. Plant Physiol.* 165, 104–113. doi: 10.1016/j.jplph.2007.05.007
- Castagnone-Sereno, P. (2006). Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. *Heredity* 96, 282–289. doi: 10.1038/sj.hdy.6800794
- Castagnone-Sereno, P., Wajnberg, E., Bongiovanni, M., Leroy, F., and Dalmasso, A. (1994). Genetic variation in *Meloidogyne incognita* virulence against the tomato Mi resistance gene: evidence from isofemale line selection studies. *Theor. Appl. Genet.* 88, 749–753. doi: 10.1007/BF01253980
- Chookajorn, T., Dzikowski, R., Frank, M., Li, F., Jiwani, A. Z., Hartl, D. L., et al. (2007). Epigenetic memory at malaria virulence genes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 899–902. doi: 10.1073/pnas.0609084103
- Cosseau, C., Azzi, A., Smith, K., Freitag, M., Mitta, G., and Grunau, C. (2009). Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters. *Mol. Biochem. Parasitol.* 166, 70–76. doi: 10.1016/j.molbiopara.2009.02.015
- Danchin, E. G., Rosso, M. N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B., et al. (2010). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17651–17656. doi: 10.1073/pnas.1008486107
- Dernburg, A. F. (2001). Here, there, and everywhere: kinetochore function on holocentric chromosomes. *J. Cell Biol.* 153, F33–F38. doi: 10.1083/jcb.153.6.F33
- Du, Z., Li, H., Wei, Q., Zhao, X., Wang, C., Zhu, Q., et al. (2013). Genome-wide analysis of histone modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in *Oryza sativa* L. Japonica. *Mol. Plant.* 6, 1463–1472. doi: 10.1093/mp/sst018
- Elango, N., Hunt, B. G., Goodman, M. A., and Yi, S. V. (2009). DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11206–11211. doi: 10.1073/pnas.0900301106
- Fneich, S., Dheilly, N., Adema, C., Rognon, A., Reichelt, M., Bulla, J., et al. (2013). 5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of *Biophalaría glabrata*, a snail intermediate host of *Schistosoma mansoni*. *Parasit. Vectors* 6:167. doi: 10.1186/1756-3305-6-167
- Freitas-Junior, L. H., Hernandez-Rivas, R., Ralph, S. A., Montiel-Condado, D., Ruvalcaba-Salazar, O. K., Rojas-Meza, A. P., et al. (2005). Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* 121, 25–36. doi: 10.1016/j.cell.2005.01.037
- Gao, F., Liu, X., Wu, X. P., Wang, X. L., Gong, D., Lu, H., et al. (2012). Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*. *Genome Biol.* 13, R100. doi: 10.1186/gb-2012-13-10-r100
- Gao, F., Wang, R., and Liu, M. (2014). *Trichinella spiralis*, potential model nematode for epigenetics and its implication in metazoan parasitism. *Front. Physiol.* 4:410. doi: 10.3389/fphys.2013.00410
- Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721
- Gómez-Díaz, E., Jordà, M., Peinado, M. A., and Rivero, A. (2012). Epigenetics of host-pathogen interactions: the road ahead and the road behind. *PLoS Pathog.* 8:e1003007. doi: 10.1371/journal.ppat.1003007
- Gutierrez, A., and Sommer, R. J. (2004). Evolution of *dnmt-2* and *mbd-2*-like genes in the free-living nematodes *Pristionchus pacificus*, *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Nucleic Acids Res.* 32, 6388–6396. doi: 10.1093/nar/gkh982
- Hendrich, B., and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* 19, 269–277. doi: 10.1016/S0168-9525(03)00080-5
- Jabbari, K., and Bernardi, G. (2004). Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* 333, 143–149. doi: 10.1016/j.gene.2004.02.043
- Jablonka, E., and Lamb, M. J. (2002). The changing concept of epigenetics. *Ann. N. Y. Acad. Sci.* 981, 82–96. doi: 10.1111/j.1749-6632.2002.tb04913.x
- Jack, A. P., Bussemer, S., Hahn, M., Pünzeler, S., Snyder, M., Wells, M., et al. (2013). H3K56me3 is a novel, conserved heterochromatic mark that largely but not completely overlaps with H3K9me3 in both regulation and localization. *PLoS ONE* 8:e51765. doi: 10.1371/journal.pone.0051765
- Jaouannet, M., Perfus-Barbeoch, L., Deleury, E., Magliano, M., Engler, G., Vieira, P., et al. (2012). A root-knot nematode-secreted protein is injected into giant cells and targeted to the nuclei. *New Phytol.* 194, 924–931. doi: 10.1111/j.1469-8137.2012.04164.x
- Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. *Science* 293, 1074–1080. doi: 10.1126/science.1063127
- Jiang, L., Mu, J., Zhang, Q., Ni, T., Srinivasan, P., Rayavara, K., et al. (2013). PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature* 499, 223–227. doi: 10.1038/nature12361
- Kornberg, R. D., and Thomas, J. O. (1974). Chromatin structure; oligomers of the histones. *Science* 184, 865–868. doi: 10.1126/science.184.4139.865
- Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J., et al. (1994). The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* 22, 1–10. doi: 10.1093/nar/22.1.1
- Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development* 130, 5083–5090. doi: 10.1242/dev.00716
- Lancelot, J., Caby, S., Dubois-Abdesselem, F., Vanderstraete, M., Trolet, J., Oliveira, G., et al. (2013). *Schistosoma mansoni* Sirtuins: characterization and

- potential as chemotherapeutic targets. *PLoS Negl. Trop. Dis.* 7:e2428. doi: 10.1371/journal.pntd.0002428
- Lepesant, J. M., Cosseau, C., Boissier, J., Freitag, M., Portela, J., Climent, D., et al. (2012). Chromatin structural changes around satellite repeats on the female sex chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence. *Genome Biol.* 13, R14. doi: 10.1186/gb-2012-13-2-r14
- Lepesant, J. M., Grunau, C., and Cosseau, C. (2011). Towards an understanding of the epigenetics of schistosomes: a comparative epigenomic study. *Mem. Inst. Oswaldo Cruz.* 106, 823–830. doi: 10.1590/S0074-02762012000100022
- Li, B., Carey, M., and Workman, J. L. (2007). The role of chromatin during transcription. *Cell* 128, 707–719. doi: 10.1016/j.cell.2007.01.015
- Luger, K., Dechassa, M. L., and Tremethick, D. J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* 13, 436–447. doi: 10.1038/nrm3382
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260. doi: 10.1038/38444
- Lyko, F., Ramsahoye, B. H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* 408, 538–540. doi: 10.1038/35046205
- Malmquist, N. A., Moss, T. A., Mecheri, S., Scherf, A., and Fuchter, M. J. (2012). Small-molecule histone methyltransferase inhibitors display rapid antimalarial activity against all blood stage forms in *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16708–16713. doi: 10.1073/pnas.1205414109
- Marek, M., Kannan, S., Hauser, A. T., Moraes Mourão, M., Caby, S., Cura, V., et al. (2013). Structural basis for the inhibition of histone deacetylase 8 (HDAC8), a key epigenetic player in the blood fluke *Schistosoma mansoni*. *PLoS Pathog.* 9:e1003645. doi: 10.1371/journal.ppat.1003645
- Mariño-Ramírez, L., Jordan, I. K., and Landsman, D. (2006). Multiple independent evolutionary solutions to core histone gene regulation. *Genome Biol.* 7:R122. doi: 10.1186/gb-2006-7-12-r122
- Melters, D. P., Paliulis, L. V., Korf, I. F., and Chan, S. W. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Res.* 20, 579–593. doi: 10.1007/s10577-012-9292-1
- Mourão, M. M., Grunau, C., LoVerde, P. T., Jones, M. K., and Oliveira, G. (2012). Recent advances in *Schistosoma* genomics. *Parasite Immunol.* 34, 151–162. doi: 10.1111/j.1365-3024.2011.01349.x
- Niu, W., Lu, Z. J., Zhong, M., Sarov, M., Murray, J. I., Brdlik, C. M., et al. (2011). Diverse transcription factor binding features revealed by genome-wide ChIP-seq in *C. elegans*. *Genome Res.* 21, 245–254. doi: 10.1101/gr.114587.110
- Perrin, C., Lepesant, J. M., Roger, E., Duval, D., Fneich, S., Thuillier, V., et al. (2013). *Schistosoma mansoni* mucin gene (SmPoMuc) expression: epigenetic control to shape adaptation to a new host. *PLoS Pathog.* 9:e1003571. doi: 10.1371/journal.ppat.1003571
- Pierce, R. J., Dubois-Abdesselem, F., Lancelot, J., Andrade, L., and Oliveira, G. (2012). Targeting schistosome histone modifying enzymes for drug development. *Curr. Pharm. Des.* 18, 3567–3578. doi: 10.2174/138161212801327248
- Rando, O. J. (2011). Genome-wide measurement of histone H3 replacement dynamics in yeast. *Methods Mol. Biol.* 759, 41–60. doi: 10.1007/978-1-61779-173-4_3
- Rosso, M. N., Favery, B., Pottie, C., Arthaud, L., De Boer, J. M., Hussey, R. S., et al. (1999). Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. *Mol. Plant-Microbe Interact.* 12, 585–591. doi: 10.1094/MPMI.1999.12.7.585
- Semblat, J. P., Bongiovanni, M., Wajnberg, E., Dalmasso, A., Abad, P., and Castagnone-Sereno, P. (2000). Virulence and molecular diversity of parthenogenetic root-knot nematodes, *Meloidogyne* spp. *Heredity* 84, 81–89. doi: 10.1046/j.1365-2540.2000.00633.x
- Simpson, V. J., Johnson, T. E., and Hammen, R. F. (1986). *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res.* 14, 6711–6719. doi: 10.1093/nar/14.16.6711
- Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45. doi: 10.1038/47412
- Subirana, J. A., and Messegue, X. (2013). A satellite explosion in the genome of holocentric nematodes. *PLoS ONE* 8:e62221. doi: 10.1371/journal.pone.0062221
- Suzuki, M. M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 9, 465–476. doi: 10.1038/nrg2341
- Triantaphyllou, A. C. (1981). Oogenesis and the Chromosomes of the Parthenogenetic Root-knot Nematode *Meloidogyne incognita*. *J. Nematol.* 13, 95–104.
- Triantaphyllou, A. C. (1983). “Cytogenetic aspects of nematode evolution,” in *Concepts in Nematode Systematics*, eds A. R. Stone, H. M. Platt, and L. F. Khalil (London: Academic Press), 55–71.
- Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* 28, 1–13. doi: 10.1016/j.molcel.2007.09.011
- Turner, B. M. (1993). Decoding the nucleosome. *Cell* 75, 5–8. doi: 10.1016/0092-8674(93)90673-E
- Turner, B. M. (2005). Reading signals on the nucleosome with a new nomenclature for modified histones. *Nat. Struct. Mol. Biol.* 12, 110–112. doi: 10.1038/nsmb0205-110
- Turner, B. M. (2012). The adjustable nucleosome: an epigenetic signaling module. *Trends Genet.* 28, 436–444. doi: 10.1016/j.tig.2012.04.003
- Urieli-Shoval, S., Gruenbaum, Y., Sedat, J., and Razin, A. (1982). The absence of detectable methylated bases in *Drosophila melanogaster* DNA. *FEBS Lett.* 146, 148–152. doi: 10.1016/0014-5793(82)80723-0
- Verhoeven, K. J., and Preite, V. (2014). Epigenetic variation in asexually reproducing organisms. *Evolution* 68, 644–655. doi: 10.1111/evo.12320
- Wenzel, D., Palladino, F., and Jedrusik-Bode, M. (2011). Epigenetics in *C. elegans*: facts and challenges. *Genesis* 49, 647–661. doi: 10.1002/dvg.20762
- Winnepenninckx, B., Backeljau, T., and De Wachter, R. (1993). Extraction of high molecular weight DNA from molluscs. *Trends Genet.* 9, 407. doi: 10.1016/0168-9525(93)90102-N
- Youngson, N. A., and Whitelaw, E. (2008). Transgenerational epigenetic effects. *Annu. Rev. Genomics Hum. Genet.* 9, 233–257. doi: 10.1146/annurev.genom.9.081307.164445

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 December 2013; accepted: 15 May 2014; published online: 06 June 2014.
Citation: Perfus-Barbeoch L, Castagnone-Sereno P, Reichelt M, Fneich S, Roquis D, Pratz L, Cosseau C, Grunau C and Abad P (2014) Elucidating the molecular bases of epigenetic inheritance in non-model invertebrates: the case of the root-knot nematode *Meloidogyne incognita*. *Front. Physiol.* 5:211. doi: 10.3389/fphys.2014.00211
This article was submitted to *Invertebrate Physiology*, a section of the journal *Frontiers in Physiology*.
Copyright © 2014 Perfus-Barbeoch, Castagnone-Sereno, Reichelt, Fneich, Roquis, Pratz, Cosseau, Grunau and Abad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

IX.3. ANNEXE 3 : PROTOCOLES DE CHIP TESTÉS SUR *P. DAMICORNIS*

Cette annexe contient un bref descriptif des différents protocoles de ChIP-Seq testés sur le corail *Pocillopora damicornis*. L'ensemble des expérimentations est résumé dans le **Tableau Supplémentaire 1**.

Tableau Supplémentaire 1 : Détail des conditions testées pour l'optimisation du ChIP, en particulier pour la digestion à la nucléase micrococcale (MNase). *CB* = Corail blanchi (réduction de > 99% des *Symbiodinium* intratissulaires); *CS* = Corail sain. *Airpic* : appareil projetant de l'air comprimé. *Waterpic* : pulvérisateur d'eau de mer.

Méthode basée sur	Échantillon	Méthodes d'extractions	Paramètres modifiés	Commentaires
<i>Cosseau & Grunau (2011)</i>	CB & CS	Airpic dans tampon de lyse	/	Peu de matériel biologique collecté Digestion sans fragments internucléosomaux
	CB & CS	Waterpic dans eau de mer	/	Peu de matériel biologique collecté
	CS	Cryobroyage à sec	/	ADN lourdement dégradé avant digestion
	CS	Cryobroyage dans tampon de lyse	/	
	CS	Cryobroyage + homogénéiseur de Dounce dans tampon de lyse	20x EDTA (2 mM final) dans tous les tampons	
	CS	Broyage manuel (mortier/pilon) dans tampon de lyse	/	
	CS	Broyage manuel (mortier/pilon) dans tampon de lyse	20x EDTA (2 mM final) dans tous les tampons	ADN moyennement dégradé
	CS	Magnalyser (abrasion par billes de céramique) dans tampon de lyse	/	
	CS	Magnalyser (abrasion par billes de céramique) dans tampon de lyse	20x EDTA (2 mM final) dans tous les tampons	Digestion sans fragments internucléosomaux
	CS	Magnalyser (abrasion par billes de céramique) + homogénéiseur de Dounce dans tampon de lyse	20x EDTA (2 mM final) dans tous les tampons Testé 0,5x, 1x, 2x, 4x de détergent dans tampon de lyse.	Digestion sans fragments internucléosomaux
	CB	Airpic dans tampon sans détergent	Osmolarité ajustée à celle de l'eau de mer	Digestion sans fragments internucléosomaux
	CB	Airpic dans tampon sans détergent + homogénéiseur de Dounce dans tampon de lyse		Digestion sans fragments internucléosomaux
	CB	Airpic dans tampon sans détergent	Osmolarité et force ionique ajustée à celles de l'eau de mer	Digestion sans fragments internucléosomaux, mais smear plus visible
	CB	Airpic dans tampon sans détergent + homogénéiseur de Dounce dans tampon de lyse		Digestion sans fragments internucléosomaux, mais smear plus visible
	CS & CB	Abrasion mécanique sur roue dans tampon de lyse	Osmolarité et force ionique ajustée à celles de l'eau de mer. 4% PVP (chélateur de polyphénols) dans tous les tampons. Digestion à 37°C et température pièce.	Digestion sans fragments internucléosomaux, mais smear plus visible et consistant sur CS
	CS	Abrasion mécanique sur roue dans tampon de lyse		Peu d'ADN récolté, en général dégradé. Digestion sans fragments internucléosomaux.
	CS	Airpic dans tampon sans détergent		
<i>Kit millipore Magna ChIP A (cat# 17-610)</i>	CB	Airpic après crosslink, dans tampon sans détergent	Sonication sur Bioraptor	ADN très endommagé, impossible d'affirmer que l'on a fragmenté autour des nucléosomes.
	CB	Airpic après crosslink, dans tampon sans détergent	Digestion MNase	Aucune digestion
<i>Schaiger et al. (2006)</i>	CB	Airpic après crosslink, dans tampon sans détergent	Digestion MNase	Aucune digestion
<i>Lefevre & Bonifer (2006)</i>	CB & CS	Airpic avant crosslink (dans eau	Digestion MNase	Peu de matériel biologique collecté
	CB & CS	Airpic après crosslink	Digestion MNase	Peu de matériel biologique collecté
	CB & CS	Abrasion mécanique sur roue dans tampon de lyse	Digestion MNase	Profil intéressant pour CS, smear de digestion avec certaines bandes plus intenses que d'autres, mais sans résultats concluant lors d'un ChIP complet

IX.4. ANNEXE 4 : DIAPOSITIVES DE SOUTENANCE

Cette annexe contient les diapositives utilisées lors de la soutenance du 8 décembre 2015. La version PowerPoint est disponible sous demande à david.roquis@yahoo.com.

David Roquis



Contribution de l'épigénétique dans les *Dauermodifikationen* et l'évolution adaptative chez le parasite humain *Schistosoma mansoni* et le corail tropical *Pocillopora damicornis*

Thèse de doctorat sous la supervision de
**Céline Cosseau
& Christoph Grunau**

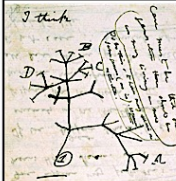
Laboratoire Interactions Hôtes-Pathogènes-Environnements
UMR5244








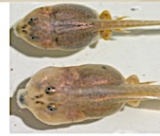



8 décembre 2015
Université de Perpignan






EXPERIMENTELLE
PROTISTENSTUDIEN
VON
VICTOR JOLLOS
I
UNTERSUCHUNGEN ÜBER VARIABILITÄT
UND VERERBUNG BEI INFUSORIEN
MIT 16 KUPFERN IM 1912
KLEINERLEBENS MIT DEM
ALBERT L. FÜR PROTISTENSTUDIEN BUCH 41

INTRODUCTION


© Plennig et al. (2010)




Paramecium spp.

Dauermodifikation

Caractère acquis transmis par le cytoplasme durant quelques générations successives, mais **non incorporé** dans l'hérédité permanente de la souche (Jollos 1939).






D. melanogaster

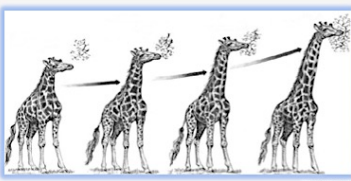


Victor Jollos 1934

Héritabilité des caractères acquis

J.B. de Lamarck C. Darwin A. Weismann



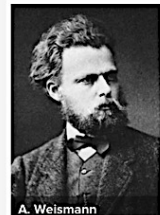


Dboudeau.fr


Introduction

4/56

Héritabilité des caractères acquis

J.B. de Lamarck C. Darwin A. Weismann






Dboudeau.fr


Introduction

4/56

Héritabilité des caractères acquis

J.B. de Lamarck C. Darwin A. Weismann



© Angela Hartsock

Introduction

4/56

Artéfacts expérimentaux?

Dauermodifikation

Caractère acquis transmis par le cytoplasme, durant quelques générations successives, mais **non incorporé** dans l'hérédité permanente de la souche (Jollos 1939).

D. melanogaster

Victor Jollos 1934

Exemples de Dauermodifikationen

Environnement normal → Présence de prédateurs

Casque

Aiguillon

Génération

Régression progressive du phénotype au fil des générations en l'absence de prédateurs

Daphnia cucullata

Daphnia spp. → Défenses inductibles en réponse aux prédateurs (Herbert & Ward 1972; Laforsch & Tollrian 1999, 2004)

Exemples de Dauermodifikationen

A. melanopus

Résistance induite à l'augmentation de température et d'acidité (Miller et al., 1999)

R. raphanistrum → Production de toxines et trichomes en réponse aux prédateurs (Agrawal et al., 1999)

H. sapiens

Une famine aux Pays-Bas induit une réduction du poids à la naissance sur trois générations (Roseboom et al. 2006, Stein et al 2007).

Nombreux exemples dans la littérature

Origines des Dauermodifikationen?

Rapides

Réversibles

Potentiellement héritables

Hors hérédité permanente

Hypothèse MÉCANISMES ÉPIGÉNÉTIQUES

Ensemble de modifications **réversibles** et **héritables**, capable d'influencer le phénotype, mais **non codées dans la séquence d'ADN** (Wu & Morris 2001).

MÉCANISMES ÉPIGÉNÉTIQUES

① Méthylation de l'ADN

Méthylation des cytosines

② Modifications post-traductionnelles des histones

③ Mécanismes basés sur l'ARN

Brins d'ARN

ADN

Nucléosome

④ Localisation nucléaire des chromosomes

Chromosome

Matouk & Marsden 2008

Modification des histones

Euchromatine

Forme relâchée permettant la transcription

Gène transcrit

Hétérochromatine

Forme condensée réprimant la transcription

Gène réprimé

Modifications des histones

Nucléosome

© janewhitney.com

Modification des histones

© millipore.com

© jannawhitey.com

Introduction

10/56

Génétique, épigénétique et phénotype

Introduction

11/56

Caractéristiques de bons modèles d'études du lien *Dauermodifikation*/Épigénétique?

Dauermodifikationen faciles à induire

Multiplication clonale

Reproduction rapide

Organismes en symbiose

Introduction

12/56

Caractéristiques de bons modèles d'études du lien *Dauermodifikation*/Épigénétique?

Schistosoma mansoni

Pocillopora damicornis

Introduction

12/56

Objectifs de la thèse

Environnement

Mécanismes moléculaires

Héritabilité

Reproduction sexuée

Multiplication asexuée / Mitose

I. Comment l'environnement influence-t-il la structure chromatinienne, les modifications d'histones et la méthylation de l'ADN? Ces changements sont-ils ciblés ou aléatoires?

II. Les modifications sont-elles héréditaires par mitose, multiplication asexuée et/ou reproduction sexuée?

Introduction

13/56

ÉTUDE DES LIENS DAUERMODIFICATIONS / ÉPIGÉNÉTIQUE

Chapitre I

Caractérisation de la structure chromatinienne chez les stades céciliaires et adultes de *S. mansoni*

Chapitre II

Modifications de la chromatine lors de l'induction de résistance à un anthelmintique chez *S. mansoni*

Chapitre III

Effet du développement dans des hôtes intermédiaires géographiquement distincts sur l'épigénome de *S. mansoni*

Chapitre IV

Développement des techniques d'étude de l'épigénome de *P. damicornis*

Rôle de la méthylation de l'ADN dans la de thermotolérance du corail *P. damicornis*

Étude de l'augmentation de la thermotolérance chez le corail tropical *Pocillopora damicornis*

Encodement de stage de M1

Rapport de stage d'Arashna Picot Picot

Article publié

Article en préparation

Article soumis

Introduction

14/56

ÉTUDE DES LIENS DAUERMODIFICATIONS / ÉPIGÉNÉTIQUE

Introduction

Chapitre I
Caractérisation de la structure chromatinienne chez les stades cercaires et adultes de *S. mansoni*

Chapitre II
Modifications de la chromatine lors de l'induction de résistance à un anthelminthique chez *S. mansoni*

Chapitre III
Effet du développement dans des hôtes intermédiaires géographiquement distincts sur l'épigénome de *S. mansoni*

Caractérisation de l'épigénome

Effet de l'environnement

Hérédité

14/56

Description du modèle *Schistosoma mansoni*

© http://galleryhip.com

Schistosoma mansoni



✓ Responsable de la schistosomiase intestinale, 67 millions de personnes infectées.
✓ Cycle de vie alternant reproduction sexuée et asexuée.
✓ Grandes capacités d'adaptations.

16/56

Cycle développemental de *S. mansoni*

Environnement: Hôte mammifère
Vers adultes mâles & femelles
M. musculus

Environnement: eau douce
Cercaires mâles & femelles
Miracidia mâles & femelles
B. glabrata
Sporocystes mâles & femelles
Environnement: Hôte mollusque

S. mansoni

17/56

Cycle de *S. mansoni*

Objectifs

- ✓ Mieux cartographier le langage des histones.
- ✓ Explorer l'influence de l'environnement sur la structure chromatinienne.
- ✓ Observer l'hérédité des différences.

Pourquoi étudier les modifications d'histones?

- ✓ Phénotypes variés au sein d'une souche avec peu de diversité génétique.
- ✓ Peu de méthylation des cytosines (Geyer *et al.* 2011, Raddatz *et al.* 2013).

S. mansoni

17/56

ÉTUDE DES LIENS DAUERMODIFICATIONS / ÉPIGÉNÉTIQUE

Chapitre I
Caractérisation de la structure chromatinienne chez les stades cercaires et adultes de *S. mansoni*

Chapitre II
Modifications de la chromatine lors de l'induction de résistance à un anthelminthique chez *S. mansoni*

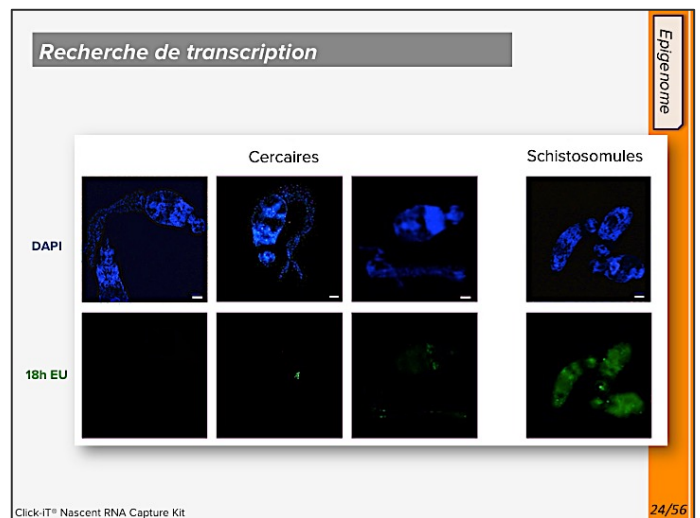
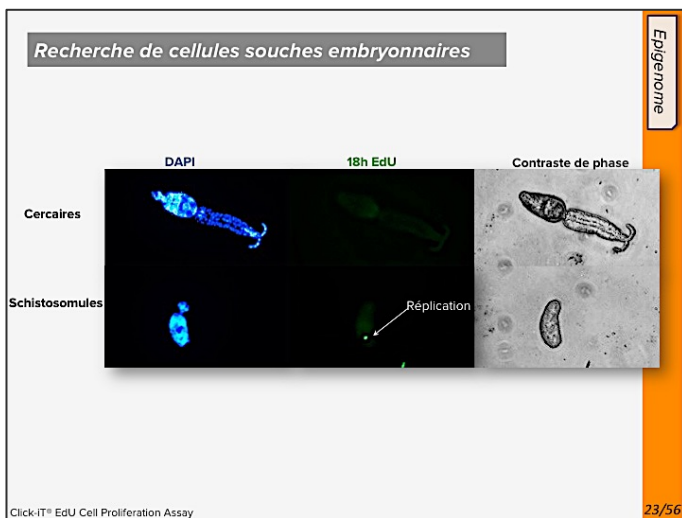
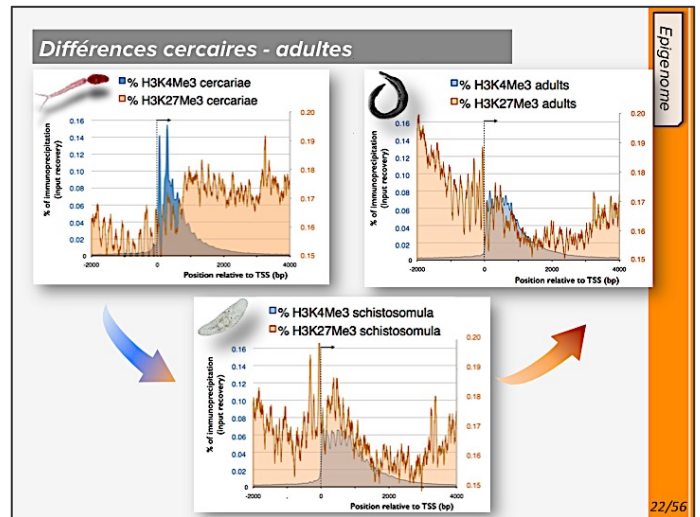
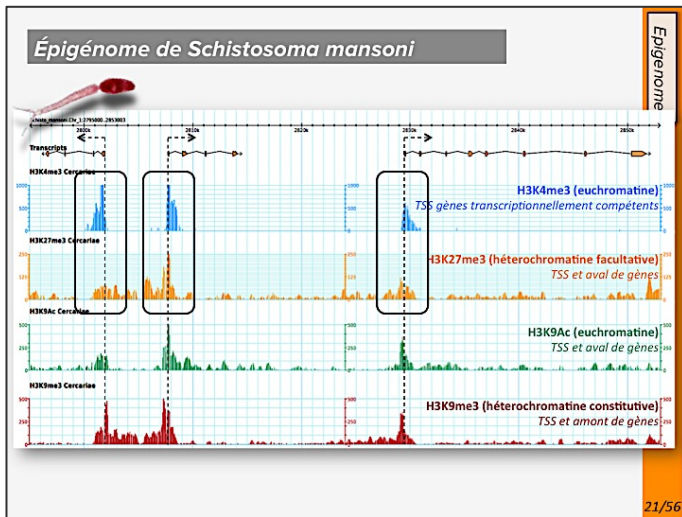
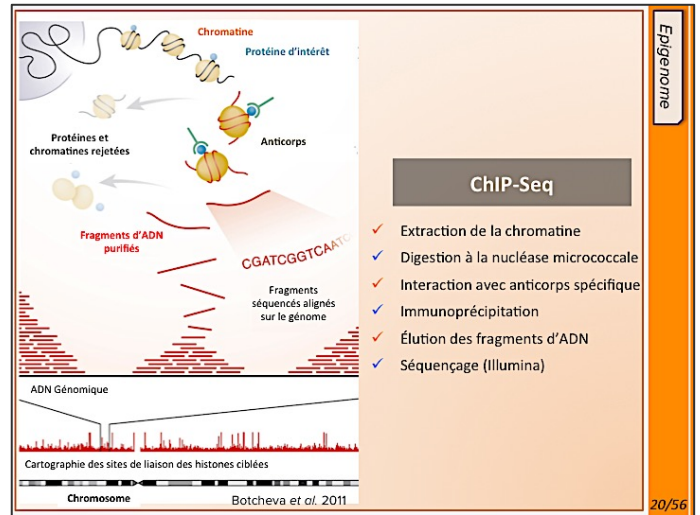
Chapitre III
Effet du développement dans des hôtes intermédiaires géographiquement distincts sur l'épigénome de *S. mansoni*

Caractérisation de l'épigénome

Effet de l'environnement

Hérédité

18/56



Résultats

- ✓ Première description de 4 modifications d'histones sur cercaires et adultes, 2 modifications sur schistosomules.
- ✓ Combinaison de marques au TSS des cercaires associée à une suspension de la transcription.

PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

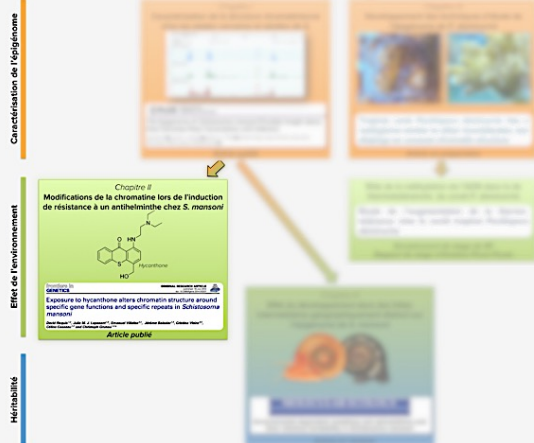
The Epigenome of *Schistosoma mansoni* Provides Insight about How Cercariae Poise Transcription until Infection

David Roquie^{1,2*}, Julie M. J. Lepesant^{1,2*}, Marion A. L. Picard^{1,2*}, Michael Freitag³, Hugues Parrinello⁴, Marco Groth⁵, Rémi Emans^{1,2}, Céline Cosseau^{1,2}, Christoph Grunau^{1,2*}

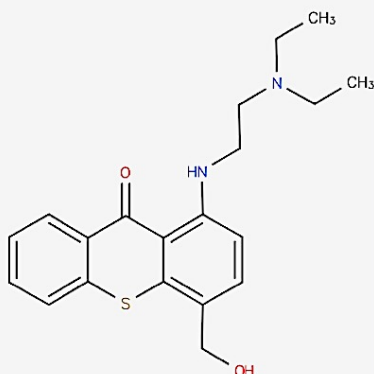
Epigénome

25/56

ÉTUDE DES LIENS DAUERMODIFICATIONS / ÉPIGÉNÉTIQUE



26/56



Effet de l'environnement #1

TRAITEMENT À L'HYCANTHONE

Environnement I

29/56

Induction de résistance à l'hycanthone

The American Journal of Tropical Medicine and Hygiene

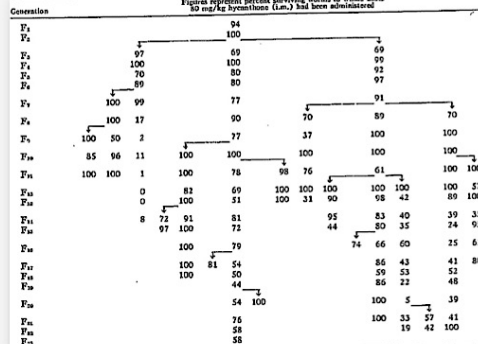
Copyright © 1973 by The American Society of Tropical Medicine and Hygiene

Vol. 10, No. 1

Printed in U.S.A.

EXPERIMENTALLY PRODUCED RESISTANCE OF SCHISTOSOMA MANSONI TO HYCANTHONE*

W. B. JANINA, S. H. ROGERS, C. L. LIU, and E. BOEDING



Environnement I

28/56

Induction de résistance à l'hycanthone

Rev. Inst. Med. trop. São Paulo
28(5):352-357, setembro-outubro, 1986

FAILURE AT INDUCING RESISTANCE TO SCHISTOSOMICIDAL DRUGS IN A BRAZILIAN HUMAN STRAIN OF SCHISTOSOMA MANSONI (1)

Luiz Cândido de Souza DIAS (2) & Celso Eduardo OLAVIER (3)

Published Online November 21 2013
Science 13 December 2013:
Vol. 342 no. 6164 pp. 1385-1389
DOI: 10.1126/science.1243106

< Prev | Table of Contents | Next >
Read Full Text to Comment (0)

REPORT

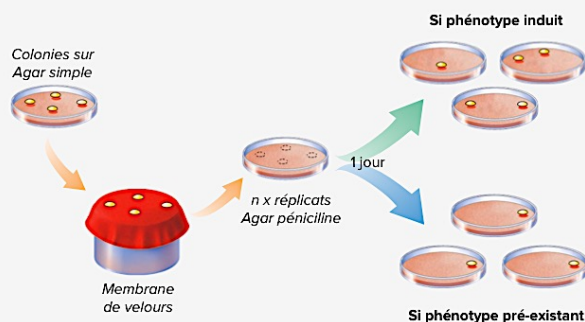
Genetic and Molecular Basis of Drug Resistance and Species-Specific Drug Action in Schistosome Parasites

Claudia L. L. Valentim^{1,2}, Donato Ciolli³, Frédéric D. Chevalier², Xiaohang Cao¹, Alexander B. Taylor¹, Stephen P. Holloway¹, Livia Pica-Mattoccia³, Alessandra Guidi³, Annalisa Basso³, Isheng J. Tsai⁴, Matthew Berriman⁴, Claudia Carvalho-Queiroz¹, Marcio Almeida², Hector Aguilar⁵, Doug E. Frantz⁵, John Hart^{1,6*}, Philip T. LoVerde^{1,2,7}, Timothy J. C. Anderson^{1,2,7}

Environnement I

29/56

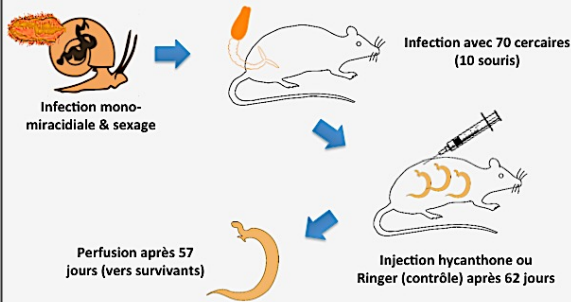
Expérience de Lederberg & Lederberg



Environnement I

30/56

Expérience hycanthone

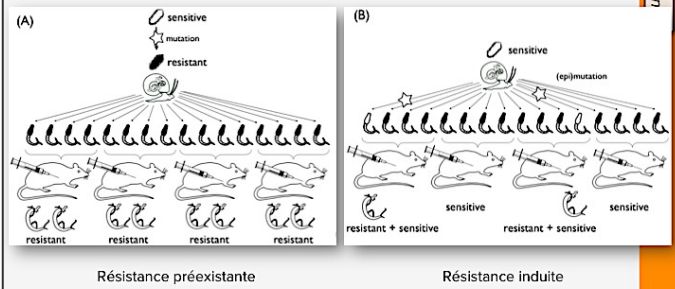


Souche du parasite présentant une proportion d'individus résistants en l'absence de la composante génétique précédemment identifiée.

Environnement I

31/56

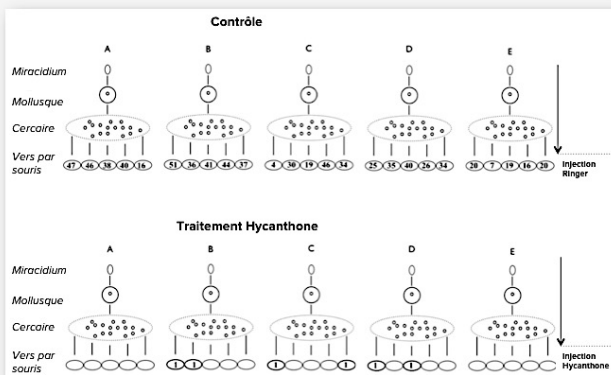
Expérience hycanthone



Environnement I

32/56

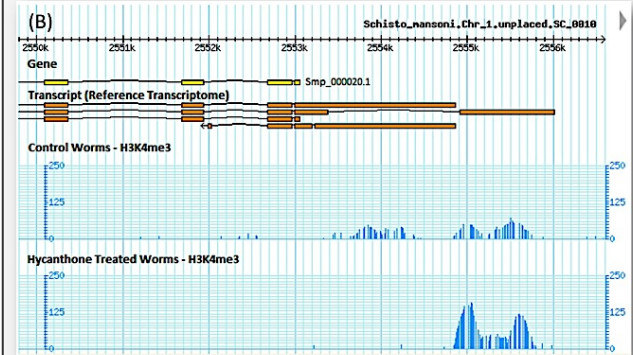
Expérience hycanthone



Environnement I

33/56

Expérience hycanthone



Environnement I

34/56

Résultats

- ✓ 69 différences de structure chromatinienne (55 dans gènes).
- ✓ Pas de voie métabolique structurante.
- ✓ Enrichissement en gènes impliqués dans l'activité catalytique (26), activité transférase (13), activité kinase (8), réponse au stress (5).
- ✓ Conclusion: inclure un suivi épigénétique pour l'apparition de résistances.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE published: 16 July 2014 doi: 10.3389/fgen.2014.00077

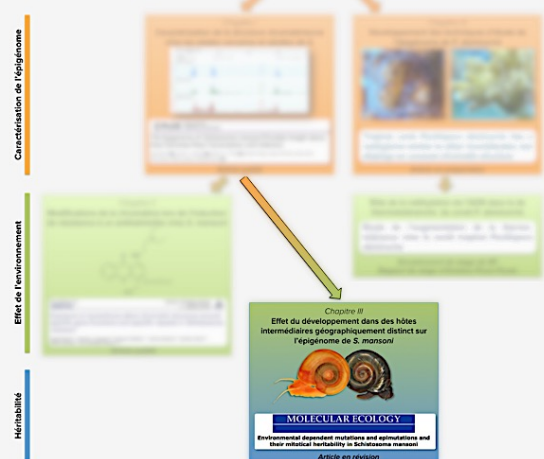
Exposure to hycanthone alters chromatin structure around specific gene functions and specific repeats in *Schistosoma mansoni*

David Roquis^{1,2}, Julie M. J. Lepesant^{1,2}, Emanuel Villafan^{1,4}, Jérôme Boissier^{1,4}, Cristina Vieira^{1,5}, Céline Cosseau^{1,2} and Christoph Grunau^{1,2*}

Environnement I

35/56

ÉTUDE DES LIENS DAUERMODIFICATIONS / ÉPIGÉNÉTIQUE



Environnement I

36/56



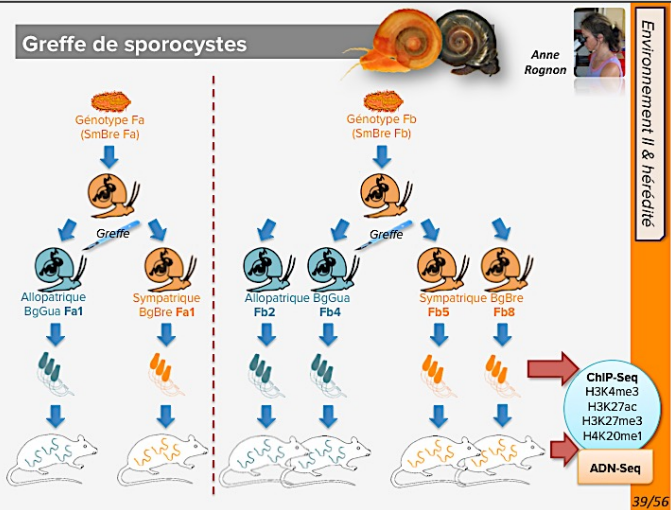
Effet de l'hôte sur le parasite



Environnement II & héritabilité

38/56

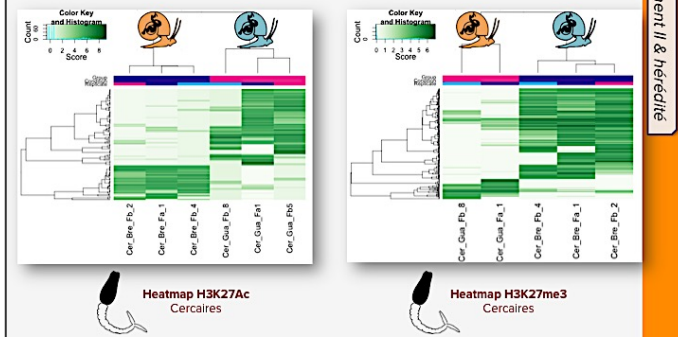
Greffe de sporocystes



Environnement II & héritabilité

39/56

Results - Diffbind

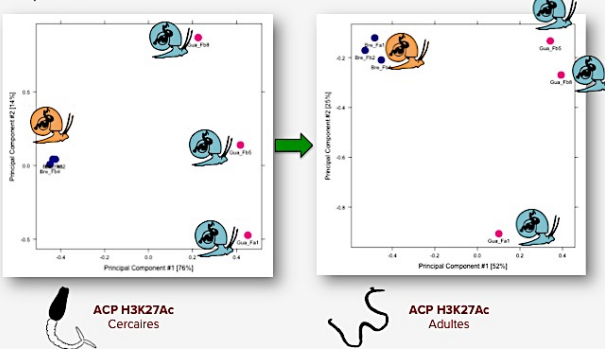


Environnement II & héritabilité

40/56

Results - Diffbind

La condition allopatrique induit de la variabilité parmi les réplicats.

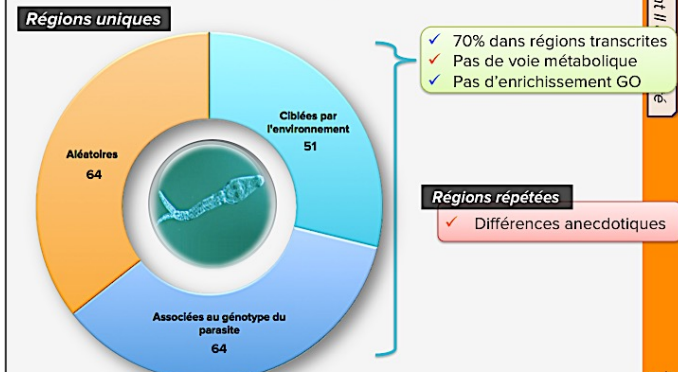


Environnement II & héritabilité

41/56

Trois types d'épimutations chez cercaires

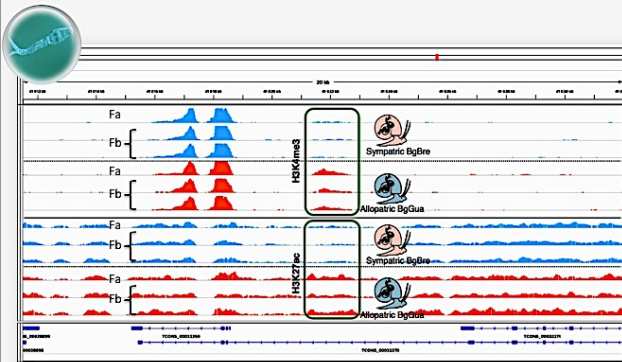
Régions où au moins deux marques d'histones sont modifiées



Environnement II & héritabilité

42/56

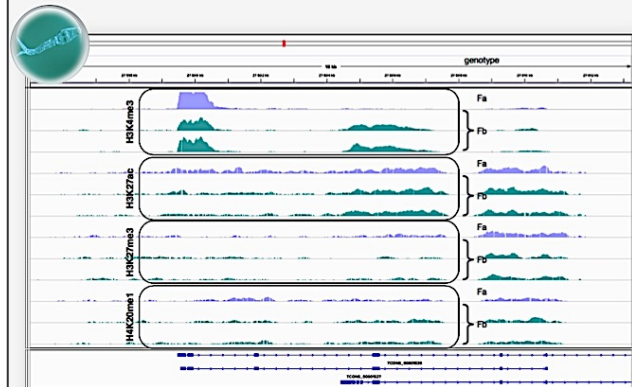
Différences ciblées par l'environnement



Environnement II & hérédité

43/56

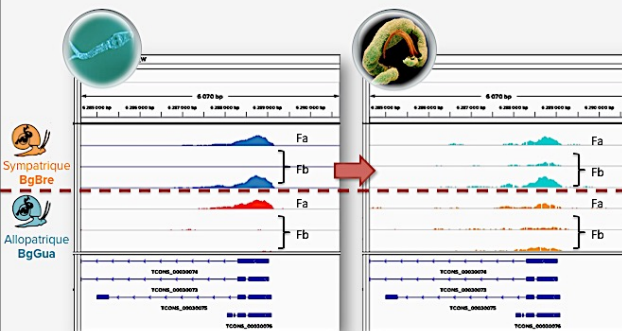
Différences associées au génotype



Environnement II & hérédité

44/56

Différences alléatoires

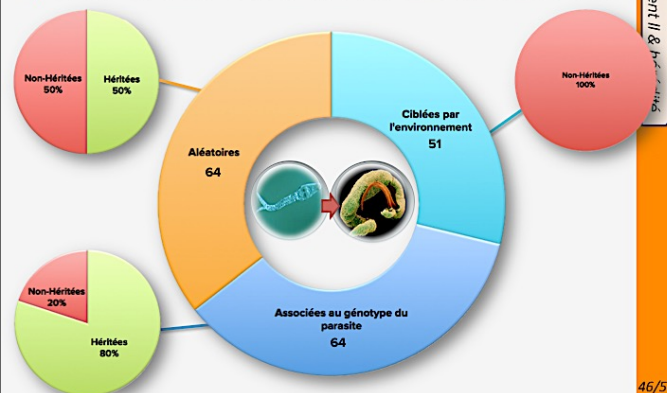


Environnement II & hérédité

45/56

Trois types d'épimutations

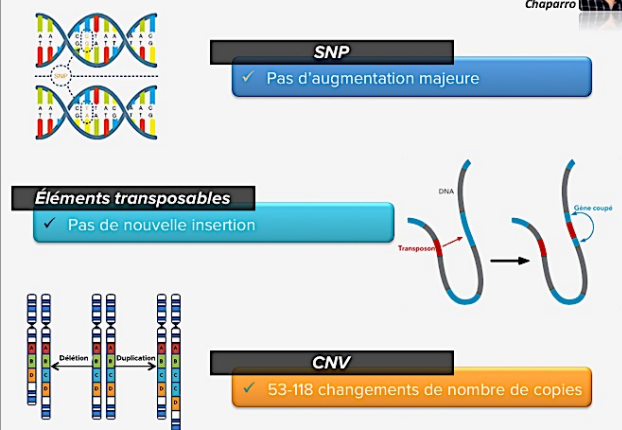
Épimutations conservées du stade cercaire au stade adulte



Environnement II & hérédité

46/56

Différences génétiques entre conditions



Environnement II & hérédité

47/56

Résultats

- ✓ Trois types d'épimutations: **ciblées**, associées au **génotype** et **aléatoires**
- ✓ Épimutations ciblées **NON-héréditaires**
- ✓ **50%** des épimutations aléatoires sont héréditaires aux adultes
- ✓ Pas d'insertion d'éléments transposables, mais augmentation des **CNV**

Accepté

MOLECULAR ECOLOGY

Environmental dependent mutations and epimutations and their mitotical heritability in *Schistosoma mansoni*

Environnement II & hérédité

48/56


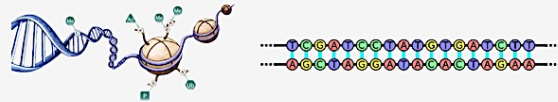


DISCUSSION

© npg.org

Considérations expérimentales




Caractéristiques d'une bonne expérience pour l'étude des implications génétiques/épigénétiques dans les Dauermodifikationen?

- I. Utilisation d'individus **génétiquement identiques**. 
- II. Analyse des modifications **épigénétiques ET génétiques**. 
- III. Analyse **individuelle** des répliquats.

Discussion

50/56

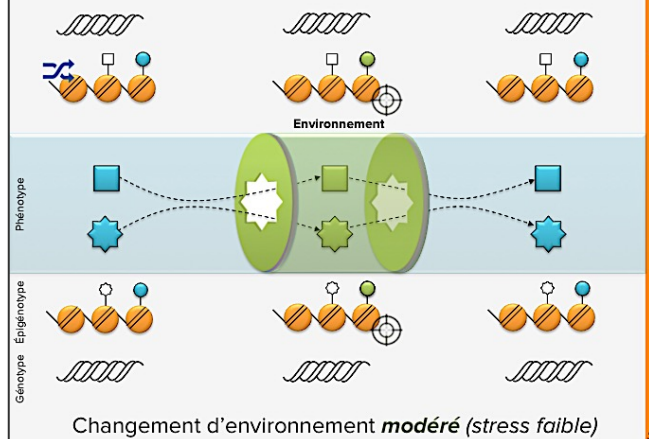
Origine des épimutations?

- I. Séquences régulatrices en cis.  Associées au génotype
- II. Induites par la condition.  Ciblées
- III. Aléatoires, mais sélectionnées par la condition.
- I. Artéfacts causés par la greffe de sporocystes.  Aléatoires
- II. Mécanisme générant de la diversité épigénétique en permanence.

Discussion

51/56

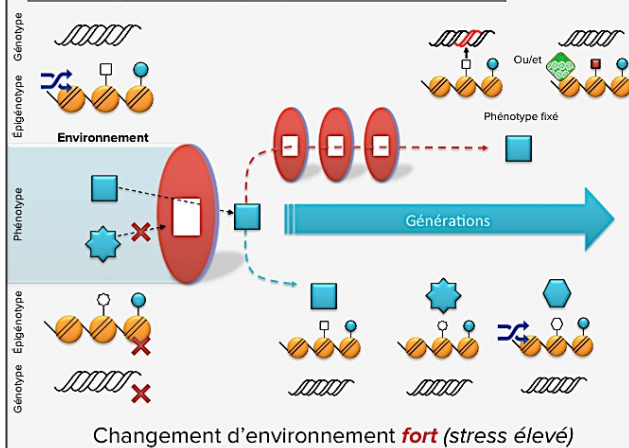
Rôle adaptatif des épimutations?



Discussion

52/56

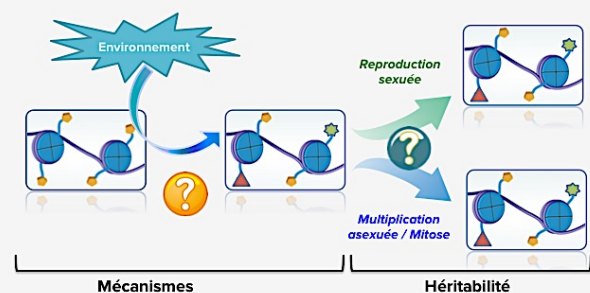
Rôle adaptatif des épimutations?



Discussion

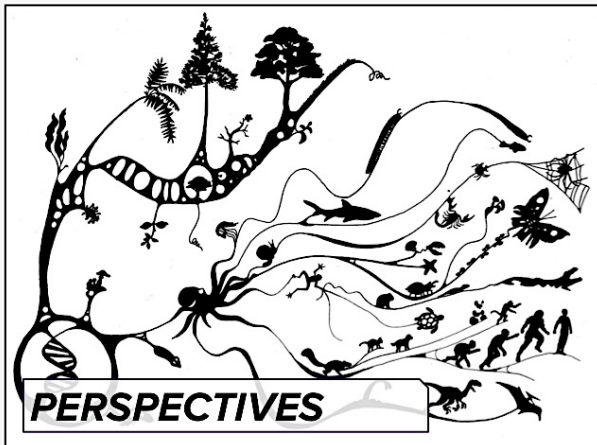
52/56

Retour sur les objectifs



- I. Comment l'environnement influence-t-il la **structure chromatinienne**, les **modifications d'histones** et la **méthylation de l'ADN**? Ces changements sont-ils ciblés ou aléatoires?
- II. Les modifications sont-elles **héréditaires** par **mitose**, **multiplication asexuée** et/ou **reproduction sexuée**?

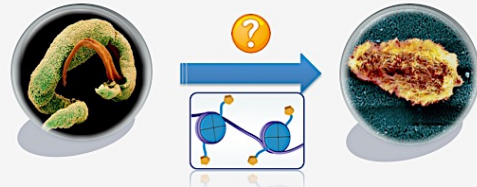
53/56



© Reserhisto

Perspectives

- ✓ Plus de répliquats pour mieux comprendre la dynamique des épimutations.
- ✓ Héritabilité sexuelle?



Perspectives

55/56

Perspectives

- ✓ Plus de répliquats pour mieux comprendre la dynamique des épimutations.
- ✓ Héritabilité sexuelle?
- ✓ Établir un lien fonctionnel avec les épimutations.

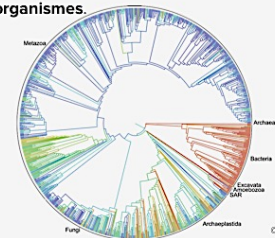


Perspectives

55/56

Perspectives

- ✓ Plus de répliquats pour mieux comprendre la dynamique des épimutations.
- ✓ Héritabilité sexuelle?
- ✓ Établir un lien fonctionnel avec les épimutations.
- ✓ Explorer d'autres organismes.



© tree.opentreeoflife.org

Perspectives

55/56



THÈSE
Pour obtenir le grade de
Docteur

Délivrée par
UNIVERSITÉ DE PERPIGNAN VIA DOMITIA

Préparée au sein de l'école doctorale ED305

Et de l'unité de recherche UMR5244

Interactions Hôtes-Pathogènes Environnement (IHPE)

Disciplines : Biologie

Spécialité : Épigénétique

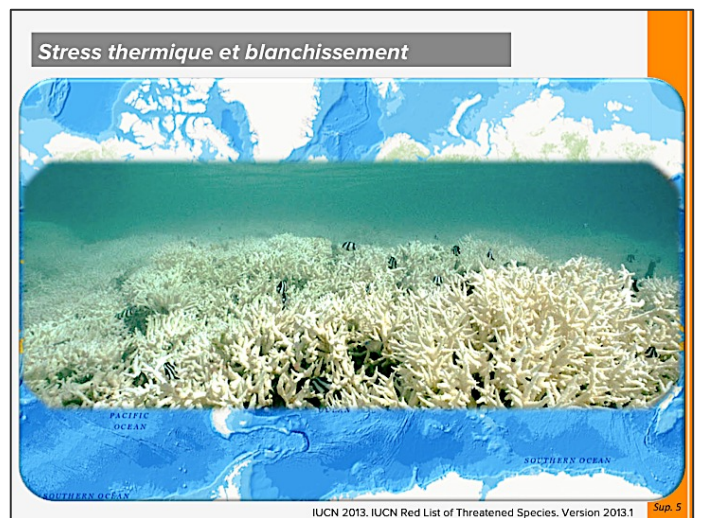
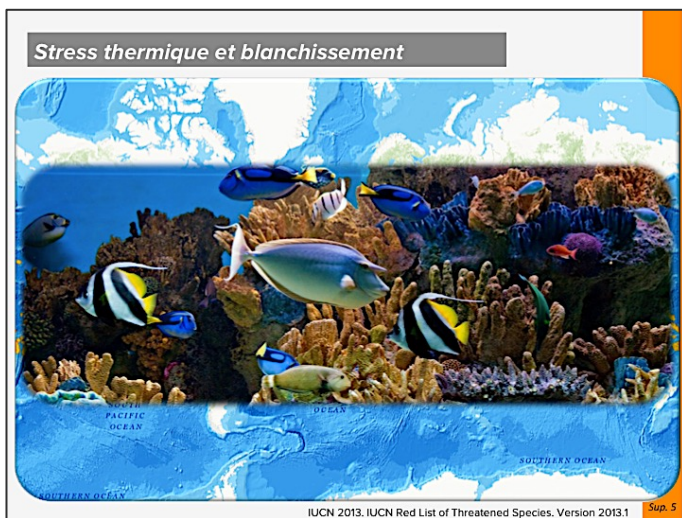
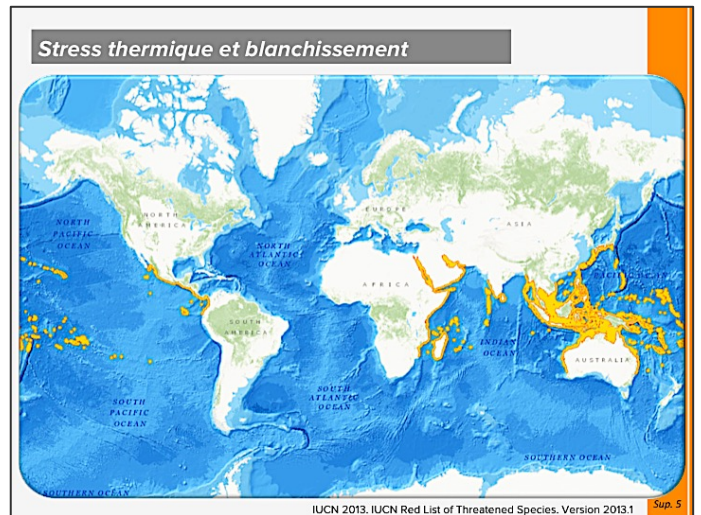
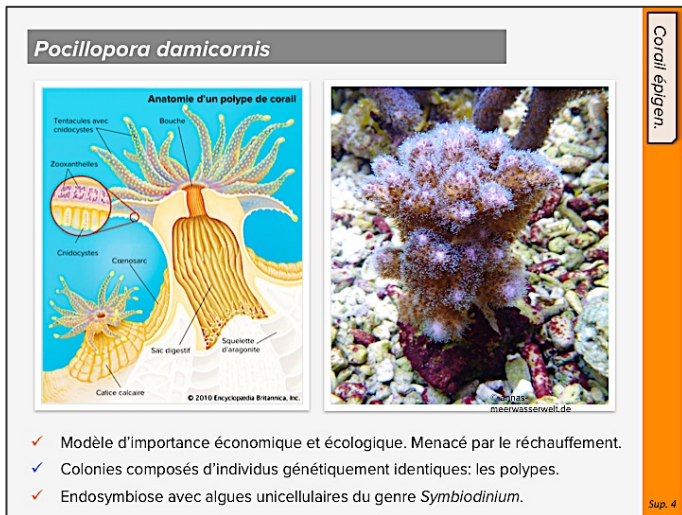
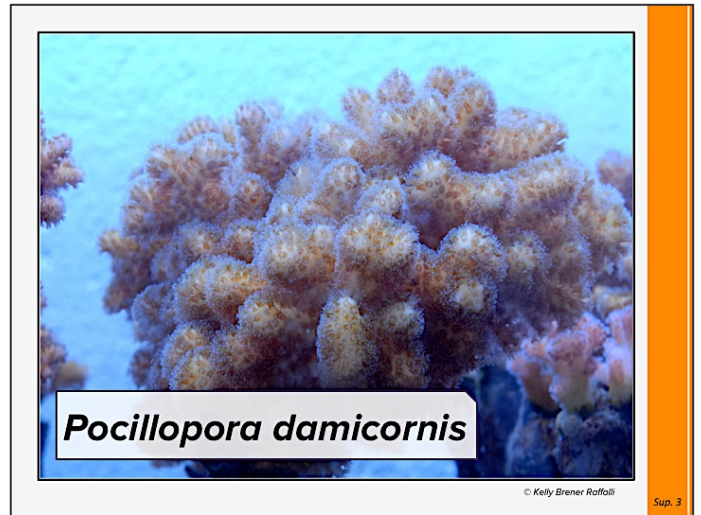
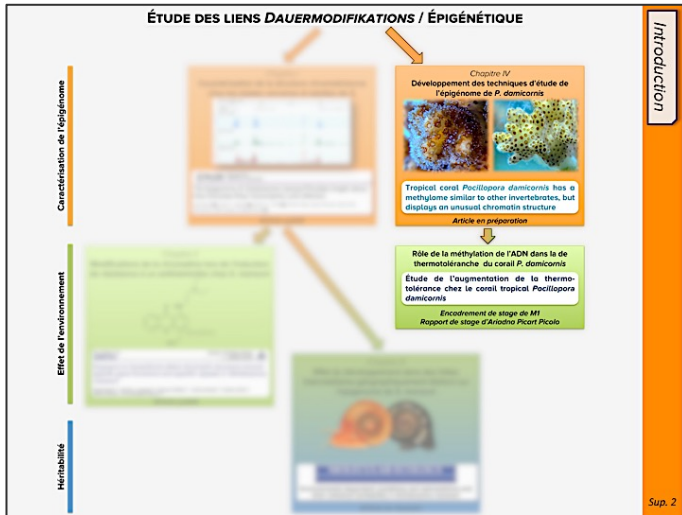
Présentée par: David ROQUIS

Contribution de l'épigénétique dans les
Dysmodifications et l'évolution adaptative chez le
parasite humain Schistosoma mansoni et le corail
tropical Pocillopora damicornis

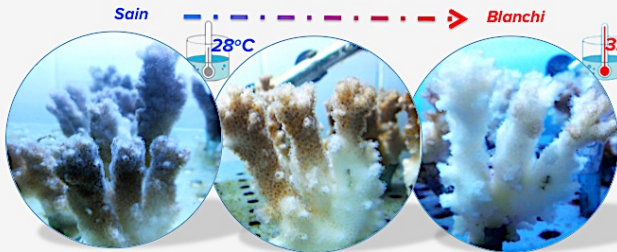
Soutenue le 9 décembre 2015 devant le jury composé de :

M. Bruno LEBLANC (PRF), Université de Perpignan

FIGURES SUPPLÉMENTAIRES



Stress thermique et blanchissement



Corail épigén.

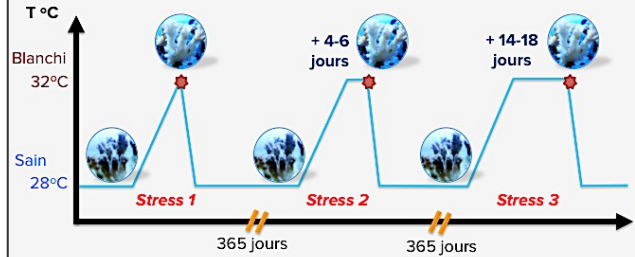
Sup. 6

Stress thermique et blanchissement

Jérémy Vidal-Dupiol



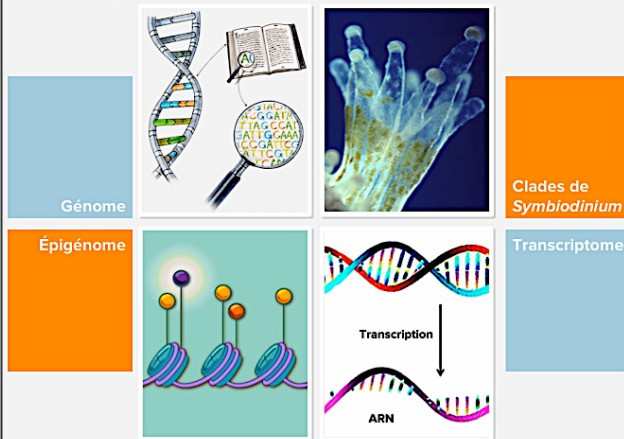
Corail épigén.



✓ La thermotolérance augmente au fil des stress.

Sup. 7

Origine(s) de la thermotolérance?



Corail épigén.

Sup. 8

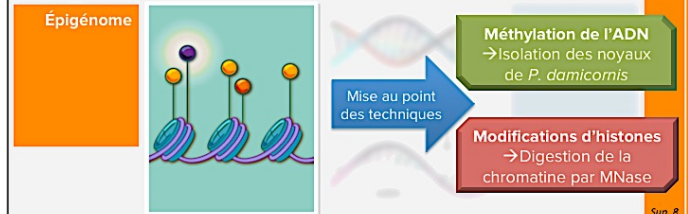
Origine(s) de la thermotolérance?

Objectifs

- ✓ Mettre au point les techniques.
- ✓ Obtenir un premier aperçu de l'épigénome du corail sain et blanchi.

Pourquoi l'épigénétique?

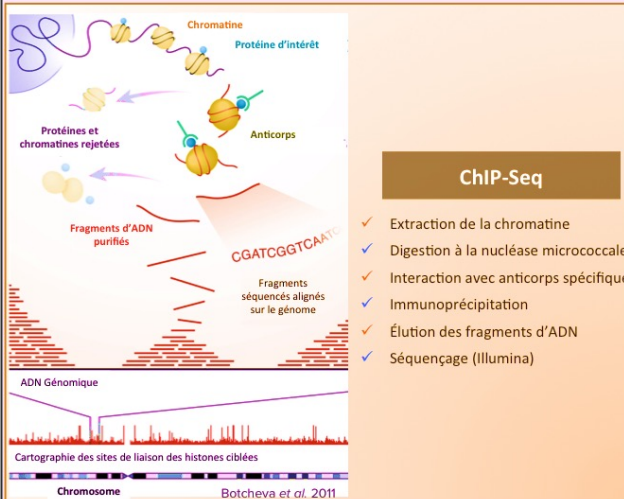
- ✓ Phénotypes de thermotolérance sur des individus clonaux.



Corail épigén.

Sup. 8

ChIP-Seq

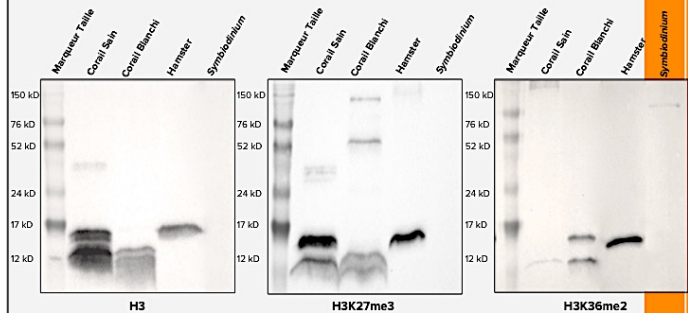


Corail épigén.

Sup. 9

Structure chromatinienne inhabituelle

Western blots sur modifications d'histones

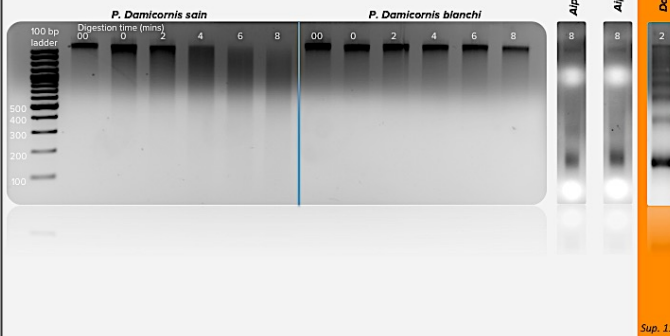


Corail épigén.

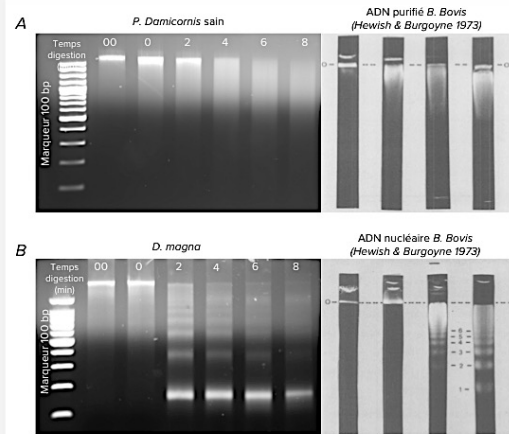
Sup. 10

Structure chromatinienne inhabituelle

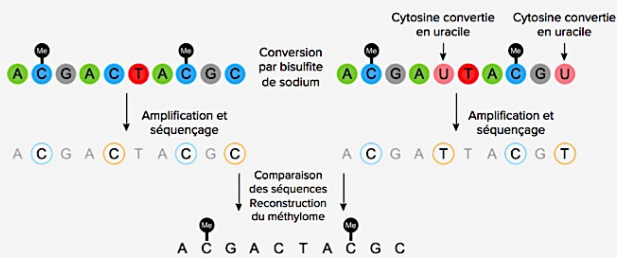
Digestion à la nucléase micrococcale



Structure chromatinienne inhabituelle



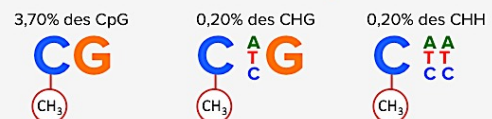
Séquençage bisulfite



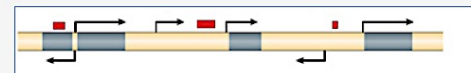
Méthylome de *P. damicornis*

✓ 0,98% de cytosines méthylées

✓ 3 contextes de méthylation

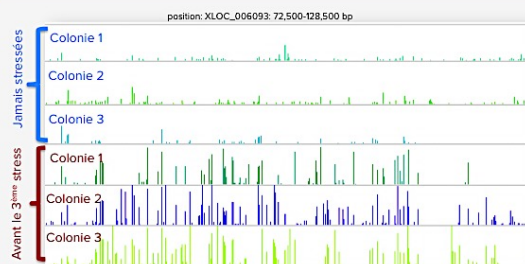


✓ Méthylation de type «mosaïque»

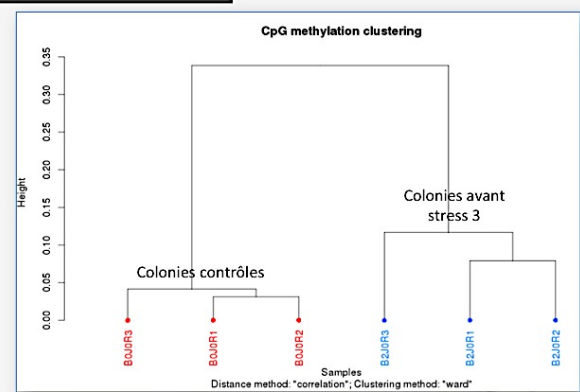


Résultats

- ✓ Structure chromatinienne inhabituelle.
- ✓ 1% de cytosines méthylées dans les trois contextes.
- ✓ Régions différenciellement méthylées entre coraux sains et blanchis → analyse en cours.



Résultats



IX.5. ANNEXE 5 : DOCUMENTS ADMINISTRATIFS RELATIFS À LA THÈSE

Cette annexe contient les documents administratifs relatifs à la thèse, présentés dans l'ordre ci-dessous :

- Avis du rapporteur Étienne Danchin sur le manuscrit de thèse (page 291).
- Avis du rapporteur Raymond Pierce sur le manuscrit de thèse (pages 293-296).
- Avis de soutenance publique (page 297).
- Procès verbal de soutenance de thèse du 8 décembre 2015 (page 299).
- Rapport sur la soutenance de thèse du 8 décembre 2015 (page 301).
- Attestation de réussite au diplôme de doctorat en biologie (page 303).

Rapport sur la thèse de David Roquis

INTITULÉE « CONTRIBUTION DE L'ÉPIGÉNÉTIQUE DANS LES *DAUERMODIFIKATIONS* ET L'ÉVOLUTION ADAPTATIVE CHEZ LES PARASITE HUMAIN *SCHISTOSOMA MANSONI* ET LE CORAIL TROPICAL *POCILLOPORA DAMICORNIS* »

en vu de la soutenance du 8 décembre 2015

J'ai lu avec grand plaisir le manuscrit de cette thèse. Le plaisir émergeait de plusieurs choses: la qualité de l'écriture, l'intérêt du sujet et une certaine forme d'érudition en profondeur.

Le préambule de 2 pages, donne le ton. Il m'a de plus fait me sentir moins seul car il décrit tellement bien mon propre parcours vis-à-vis de ce que l'on entend par hérédité. C'est en soit une information très rassurante car le cursus de David est vraiment très différent du mien, sans parler du décalage dans le temps sur lequel je n'insisterai pas.

L'introduction est magistrale à mes yeux. Elle dresse la vision de l'impétrant de l'évolution et de l'adaptation depuis les origines grecs de la science jusqu'à nos jours en un zoom avant qui donne un peu le vertige, signe de qualité. Cette section introduit la notion de *Dauermodifikation*, qui n'est rien d'autre à mes yeux que la première formulation scientifique de ce que l'on appelle aujourd'hui l'hérédité non génétique. Ce terme constitue le fil directeur de cette thèse comme en témoigne la présence de ce concept dans le titre même de la thèse. Il se peut qu'une partie de mon enthousiasme soit lié aux sections qui suivent l'introduction de ce concept. J'aurai certainement envie de discuter certains aspects de la figure 6 qui précède la section intitulée épigénétique. Là de nouveau une culture générale tout à fait remarquable et très pédagogiquement présentée. La section 1.2.2 pourrait avoir été baptisée autrement, mais tout cela reste passionnant. Je reviendrai certainement sur des questions qui ont émergé dans ma tête lors de cette lecture, en particulier à la lecture de la figure 15. La fin de cette introduction achève le zoom avant en présentant la problématique générale de la thèse.

S'ensuivent 4 chapitres, tous correspondant à des articles, certains publiés, d'autres dans le processus de soumission, d'autres quasi fini mais gardé sous le coude pour accompagner une publication en devenir. Bref une belle production scientifique, réalisée ou en devenir. L'ensemble de ces parties qui constituent le corps de la thèse m'ont fait réaliser à quel point certaines approches sont fortement dirigée, voire contraintes par les aspects purement techniques des méthodes de biologie moléculaire. Je suppose que c'est en réaction à cet aspect flagrant que David a désiré avoir une vraie section conceptuelle en amont et je ne peux que le féliciter pour ce choix qui révèle pour moi une vraie dimension de recherche. C'est d'ailleurs de l'interaction entre ces chapitres et l'introduction qu'ont émergé les questions que je poserai lors de la soutenance. N'étant pas un spécialiste des aspects purement techniques, je m'en tiendrai plus aux parties conceptuelles. Mais d'une manière générale, je trouve que cette thèse nous fait toucher du doigt à quel point nous ne connaissons que très peu de choses concernant ce processus qui va du gène à l'individu. À côté de cela la partie purement génétique apparaît comme une autoroute de savoir (même si nous connaissons tous toutes les limites de nos connaissances en génétique).

Même si j'en devine la raison, je ne peux me retenir de formuler une question qui m'a sauté aux yeux à l'abord du chapitre V: pourquoi avoir introduit, si tard, un deuxième modèle biologique, certainement passionnant, mais qui apporte une couche de complexité supplémentaire pas forcément évidente à assumer.

Arrive ensuite une dernière section consacrée à la discussion qui se révèle tout aussi intéressante.

Bien évidemment mon avis est des plus favorables pour la soutenance de cette thèse à laquelle je me réjouis donc de participer.



Etienne Danchin



Raymond Pierce
Equipe 2 « Molecular Biology of
Schistosome Development and
Reproduction »

Centre d'Infection et d'Immunité de Lille
Inserm U1019 – CNRS UMR 8204
Institut Pasteur de Lille
Univ Lille Nord de France

Tel : 33-(0)3 20 87 77 83
Fax : 33-(0)3 20 87 79 88
Mél : raymond.pierce@pasteur-lille.fr

N/Réf : DRV/JP/2015
Objet : rapport scientifique sur la thèse de David Roquis

Université de Perpignan Via Domitia
Direction de la Recherche et de la Valorisation
52, avenue Paul Alduy
66860 Perpignan Cedex 09

Lille, le 16 novembre 2015

**Rapport sur le mémoire déposé par M. David ROQUIS pour l'obtention du grade de Docteur de l'Université de
Perpignan Via Domitia**

Rapporteur : Raymond J. Pierce, CIIL, Inserm U 1019, CNRS UMR 8204, Université de Lille, Institut Pasteur de Lille.

Titre du mémoire : Contribution de l'épigénétique dans les *Dauermodifikationen* et l'évolution adaptative chez le parasite
humain *Schistosoma mansoni* et le corail tropical *Pocillopora damicornis*

Le manuscrit de M. ROQUIS se démarque des thèses habituellement présentées par une certaine liberté de ton et d'un point de vue critique, à la fois sur le thème de recherche et les résultats obtenus, tout à fait rafraîchissante. Le parcours de carrière atypique de M. ROQUIS lui permet sans doute un plus grand recul que d'habitude par rapport aux objectifs et les expériences effectuées. Par conséquent, ses discussions et conclusions sont très fournies et montrent une grande maturité de réflexion scientifique.

Le travail décrit dans le manuscrit concerne des approches à l'étude de la contribution des modifications épigénétiques à la variabilité phénotypique et l'évolution adaptative du parasite plathelminthe *Schistosoma mansoni* et du corail *Pocillopora damicornis*. Globalement, cette thèse est claire, bien présentée, très intéressante et agréable à lire. Les résultats sont nouveaux et significatifs et le travail décrit, considérable, se décline en quatre chapitres décrivant les études expérimentales, correspondant chacun à un article. Deux d'entre eux sont déjà publiés, un autre est soumis pour publication et un autre est présenté sous une forme quasi-définitive pour la soumission.

CIIL, Inserm U1019-CNRS UMR 8204
Institut Pasteur de Lille
Bâtiment Gernez-Rieux
1, rue du Pr A Calmette, BP 245 - 59019 Lille Cédex, France

Commentaires détaillés

L'introduction générale commence par une présentation des concepts de l'adaptation et de la plasticité phénotypique, notions à la base des travaux présentés. Suit une présentation historique des idées et théories de l'évolution adaptative très large, allant de la Grèce de l'Antiquité, via Lamarck et Darwin, jusqu'à l'émergence de la génétique, la « nouvelle synthèse » de Julian Huxley et le génocentrisme. Cette synthèse, bien qu'assez complet, comporte inévitablement des choix car la littérature est très vaste. Néanmoins, on peut regretter que, par exemple, dans le contexte de la sélection du comportement altruiste, Hamilton soit mentionné, mais que George Price ne l'est pas. Par la suite est abordé le sujet de l'hérédité des caractères acquis, d'abord via les travaux controversés de Jollos et les « *Dauermodifikationen* », mais qui trouvent une justification actuellement avec des études plus récentes, d'abord chez des plantes, et la mise en évidence des bases épigénétiques de cette hérédité non-mendélienne non-portée par la séquence de l'ADN. La description de l'épigénétique moderne est très complet concernant la méthylation de l'ADN et les modifications des histones. La place de la description de la méthodologie ChIP-Seq dans ce contexte est discutable et aurait pu se retrouver dans l'introduction d'un chapitre de résultats, ou en annexe. De plus, la section concernant la localisation nucléaire des chromosomes, résultant de modifications structurales de la chromatine, est pauvrement documentée en comparaison d'autres parties de l'introduction (la référence citée la plus récente date de 2010) et mériterait une présentation plus complète. Enfin, sont présentés les modèles d'étude : le parasite plathelminthe *S. mansoni* et le corail tropical *P. damicornis*. Dans les deux cas, la présentation des organismes, leur importance et leur intérêt pour l'étude est très complète.

L'objectif global de la thèse, ambitieux, était d'étudier les liens éventuels entre phénotypes induits par l'environnement, identifiés comme des *Dauermodifikationen*, et les modifications de la structure de la chromatine. Deux questions principales ont été posées :

- Est-ce que les changements de structure de la chromatine observés entre des organismes sujets aux conditions environnementales différentes sont aléatoires, ciblés ou les deux à la fois ?
- Est-ce que ces changements sont transmissibles, soit de façon asexuel (mitose) ou sexuel (méiose) ?

Pour y répondre, différentes approches ont été abordées avec les deux modèles d'étude cités plus haut et d'autres objectifs, spécifiques à chaque modèle ont été exploré. Ainsi, le premier chapitre des résultats concerne une comparaison de l'épigénome de *S. mansoni*, plus précisément la présence de quatre modifications d'histones, à deux stades différents, chez les cercaires (larve infestante) et les

vers adultes. Ces travaux, publiés dans *PLoS NTD*, ont permis de démontrer la présence simultanée de deux de ces marques, la triméthylation de H3K27 (synonyme de la répression de la transcription) et de H3K4 (synonyme de l'activation de la transcription) sur au moins 121 gènes chez les cercaires. Cette bivalence de la méthylation est retrouvée chez les cellules souches embryonnaires (ESC) et c'est la première fois qu'elle est retrouvée en dehors des vertébrés. Cette étude a également montré l'absence de transcription chez les cercaires, soutenant l'hypothèse qu'à ce stade, la transcription est en suspens (« *poised* ») avant l'invasion de l'hôte définitif. Ces résultats et cette publication sont d'une grande originalité. Un seul bémol : les fichiers supplémentaires (certes disponibles on-line) ne sont pas présentés dans la thèse et pourraient être inclus dans sa version définitive.

Dans le but de déterminer l'impact environnemental sur l'épigénome de *S. mansoni*, M. ROQUIS a d'abord étudié l'effet de l'hycanthone sur le parasite. Certains travaux montrent que ce composé, utilisé brièvement pour le traitement de la schistosomose, en dose sous-curative est capable d'induire une résistance à des doses curatives qui se transmet, à des proportions fluctuantes, à des générations de parasites suivantes en l'absence de traitement supplémentaires. Toutefois, les travaux menés au laboratoire de Perpignan ne montrent pas le même degré d'induction de la résistance. Néanmoins des expériences de ChIP-Seq ont été effectuées sur des parasites résistants et sensibles. Cette approche a permis l'identification d'au moins 64 changements de la structure chromatinienne (modifications d'histones) concernant des gènes impliqués dans des fonctions de catabolisme, de la détoxification et l'activité de transfert. Notamment, un profil différentiel de la marque H3K27me3 dans le corps du gène codant la SmMRP1, associé à la résistance aux drogues, pourrait être corrélé avec le phénotype de résistance à l'hycanthone. Toutefois, aucune modification n'a été constatée concernant le gène codant une sulfotransferase, associée à la résistance à l'hycanthone à transmission Mendélienne. Ces résultats ont été publiés dans le journal *Frontiers in Genetics*.

La troisième étude effectuée chez le parasite *S. mansoni*, concerne un autre type d'impact environnemental, cette fois fourni par l'hôte intermédiaire, le mollusque d'eau douce, *Biomphalaria glabrata*. La même souche (SmBre) du parasite a été utilisée pour l'infection (mono-miracidiale) de deux souches de mollusques sympatriques (BgBre) et allopatriques (SmGua). La technique de greffe de sporocysts a ensuite permis d'étudier d'obtenir de nombreux mollusques infectés par un même génotype de *S. mansoni* et, ainsi, d'obtenir des réplicats biologiques d'une même condition avec un fond génétique identique. La principale différence observée était une variation dans le nombre de copies (CNV) de 50 à 100 gènes. De plus, trois types de changements de l'épigénome ont été identifiés : dépendant de l'environnement, au hasard ou dépendant du génotype. Toutefois, seuls 50% des changements « au hasard » se sont avérés transmissibles au vers adultes. Ces modifications

ont le potentiel de contribuer à l'évolution adaptative, mais ne constituent pas une base pour l'héritabilité de *Dauermodifikationen*. Ces dernières n'ont donc pas encore été démontrées au niveau des modifications de la chromatine chez le parasite. Ces travaux sont soumis à *Molecular Ecology*.

Le dernier chapitre des résultats concerne l'étude de l'épigénome du corail tropical *P. damicornis*. Ce modèle, très séduisant et d'une importance capitale dans le contexte du réchauffement climatique, offre la possibilité d'étudier le rôle de l'environnement (stress thermique) sur l'épigénome et l'adaptation du corail à ce stress. Toutefois, il a fallu d'abord jeter les bases de l'étude de l'épigénome de cet organisme. Tout d'abord M. ROQUIS a développé une méthodologie permettant l'isolement des noyaux cellulaires de *P. damicornis* et leur séparation de celles de l'organisme symbiotique mutualiste *Symbiodinium*, un dinoflagellé possédant un génome dix fois plus grand que celui de *P. damicornis*. L'obtention de ces noyaux d'un degré de purification suffisante a ensuite permis la caractérisation du méthylome de ce corail. De plus, des résultats non inclus dans le manuscrit présenté (en préparation) montrent une augmentation de la méthylation de l'ADN (m5C) en réponse au stress thermique. Un autre résultat fort intéressant concerne la caractérisation de la structure de la chromatine de *P. damicornis*. Il s'avère que la structure en nucléosomes habituel trouvé chez les métazoaires, soit absente chez ce corail. Toutefois, ces résultats restent à confirmer. L'ensemble des résultats obtenus permettront l'étude approfondie des modifications épigénétiques induites par des stress environnementaux et leur héritabilité.

La discussion générale, très fournie, reprend en détail les résultats obtenus et les analyse de façon exhaustive, soulignant la capacité critique de M. ROQUIS envers de ses résultats et sa capacité d'en tirer des conclusions permettant modifier ses approches. Plutôt que de soulever ces différents points dans ce rapport, il sera beaucoup plus approprié d'en discuter lors de la soutenance.

En conclusion, le travail présenté fournit une masse impressionnante de données d'une excellente qualité, utilisant des techniques et stratégies les plus performantes. En plus des quatre articles publiés, soumis ou en préparation déjà cités, M. ROQUIS a également participé à l'élaboration de deux autres articles déjà publiés. Les quelques critiques que j'ai évoquées sont plus que contrebalancées par l'excellent niveau du travail. Je félicite M. ROQUIS et je donne un avis très favorable quant à la recevabilité du mémoire et à la soutenance de cette thèse.



Dr Raymond J. Pierce, PhD.
Head of Team 2 - CIIL
"Molecular Biology of Schistosome
Development and Reproduction"
Inserm U1019 - CNRS UMR8204
Univ Lille Nord de France
Institut Pasteur de Lille

École doctorale 305 «Énergie Environnement »

**AVIS DE PRÉSENTATION DE TRAVAUX
EN VUE DE L'OBTENTION DU DOCTORAT**

Monsieur ROQUIS David soutiendra **mardi 8décembre 2015 à 14h – à l'amphi 5** de l'Université de Perpignan Via Domitia, un doctorat spécialité : Biologie.

TITRE DE LA THÈSE:

Contribution de l'épigénétique dans les *Dauermodifikationen* et l'évolution adaptative chez le parasite humain *Schistosoma mansoni* et le corail tropical *Pocillopora damicornis*

RÉSUMÉ :

De nombreux travaux, chez divers organismes vivants, font état de caractères acquis en réponse à des changements environnementaux et pouvant être transmis à la descendance sur quelques générations, avant de disparaître progressivement. Ces phénomènes, autrefois nommés *Dauermodifikationen*, ne suivent pas les processus génétiques et évolutifs classiques. Toutefois, les mécanismes épigénétiques, plus flexibles, sont des candidats prometteurs pour expliquer ces observations. Dans le cadre de cette thèse, nous avons exploré les liens entre épigénétique et *Dauermodifikationen* chez deux organismes distincts : le corail tropical *Pocillopora damicornis* et le parasite humain *Schistosoma mansoni*. Nos travaux ont permis une meilleure caractérisation des épigénomes de ces deux organismes. Nous avons également constaté que des modifications de l'épigénome peuvent apparaître *a priori* de manière spontanée ou être induites par des changements environnementaux. Certaines de ces modifications sont transmises et persistent lors du passage entre des stades développementaux distincts.

Directeurs de thèse

M. CHRISTOPH GRUNAU, Professeur à l'Université de Perpignan Via Domitia, et Mme CÉLINE COSSEAU, Maître de conférences

Laboratoire où la thèse a été préparée : UMR 5244, IHPE – UPVD

Le jury sera composé de

Mme. CÉLINE COSSEAU	MCF – Université de Perpignan Via Domitia
M. TIENNE DANCHIN	Directeur de Recherche – Université Toulouse 3, Paul Sabatier
M. CHRISTOPH GRUNAU	Professeur – Université de Perpignan Via Domitia
Mme. MARIE MIROUZE	Chargé de Recherche – Université de Montpellier 2
M. GUILLAUME MITTA	Professeur – Université de Perpignan Via Domitia
M. RAYMOND PIERCE	Directeur de Recherche – Institut Pasteur de Lille, CIIL
M. OLIVIER REY	Chargé de Recherche – University Singleton Park – Swansea, Pays de Galles

Etudiant : M. DAVID ROQUIS né le : 24/10/1982

Diplôme : DOCTORAT - BIOLOGIE

Titre des travaux : Contribution de l'épigénétique dans les Dauermodifications et l'évolution adaptative chez le parasite humain *Schistosoma mansoni* et le corail tropical *Pocillopora damicornis*

Ecole doctorale : Energie et Environnement

Directeur : M. CHRISTOPH GRUNAU

Codirecteur : Mme CELINE COSSEAU

Lieu de soutenance : UPVD - Amphi 5

La soutenance est publique.

Résultat : *OK*

Mention : *Très honorable*

Avis de reproduction : *OK*

Membres du Jury

Nom	Qualité	Etablissement	Rôle	Signature
Mme CELINE COSSEAU	MAITRE DE CONFERENCE	UNIVERSITE DE PERPIGNAN		<i>[Signature]</i>
M. ETIENNE DANCHIN	DIRECTEUR DE RECHERCHE	UNIVERSITE TOULOUSE 3 PAUL SABATIER		<i>[Signature]</i>
M. CHRISTOPH GRUNAU	PROFESSEUR DES UNIVERSITES	UNIVERSITE DE PERPIGNAN		<i>[Signature]</i>
Mme MARIE MIROUZE	CHARGE DE RECHERCHE			<i>[Signature]</i>
M. GUILLAUME MITTA	PROFESSEUR DES UNIVERSITES	UNIVERSITE DE PERPIGNAN		
M. RAYMOND PIERCE	DIRECTEUR DE RECHERCHE CNRS			<i>[Signature]</i>
M. OLIVIER REY <i>GR</i>	CHARGE DE RECHERCHE			<i>[Signature]</i>

Autorisation de reproduction :

☒ OUI

☐ OUI après corrections

☐ Thèse confidentielle jusqu'en :

☐ NON

Rapport de soutenance :

Le président signe le rapport de soutenance qui est contresigné par l'ensemble des membres du jury.

Il doit parvenir à la DRV dans un délai d'un mois maximum après la soutenance.

RAPPORT DE SOUTENANCE DU DOCTORAT

Spécialité : BIOLOGIE

NOM – Prénom : Monsieur David ROQUIS
Date de la soutenance : Mardi 8 décembre 2015 à 14h
UPVD – Amphi 5

Sujet ou intitulé des travaux :

Contribution de l'épigénétique dans les *Dauermodifikationen* et l'évolution adaptative chez le parasite humain *Schistosoma mansoni* et le corail tropical *Pocillopora damicornis*

RAPPORT :

Le jury a tenu à souligner la qualité exceptionnelle de l'écrit de monsieur David Roquis. Ce document a révélé une maturité et une érudition rares. Lors de la soutenance, David a présenté les résultats de ses travaux de recherche dans un exposé de 45 minutes clair, didactique et remarquablement illustré. La présentation orale a révélé les qualités de chercheur du candidat, sa maîtrise du sujet, sa capacité d'anticipation et de projection. Il a su développer une argumentation pertinente en réponse aux questions posées par un jury composé de membres aux champs d'études variés. Le jury voulait également mettre en exergue dans ce rapport l'interface entre biologie évolutive (supra individuelle) et la biologie infra individuelle de cette thèse. David a su intégrer tant les concepts que des méthodes variées relevant de ces deux domaines de la biologie. Ses développements en bioinformatique ont été qualifiés de particulièrement remarquables.

A l'unanimité, le Jury a décidé d'attribuer à Monsieur David Roquis, le grade de Docteur de L'Université de Perpignan Via Domitia avec la plus haute mention (très honorable). Le jury y ajoute ses félicitations orales.

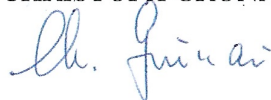
Mention : ☐ Honorable ☒ Très Honorable

Les membres du jury : (souligner le nom du président du jury) :

Mme CELINE COSSEAU



M. CHRISTOPH GRUNAU



M. GUILLAUME MITTA



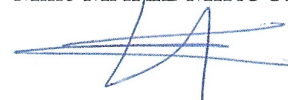
M. OLIVIER REY




M. ETIENNE DANCHIN



Mme MARIE MIROUZE



M. RAYMOND PIERCE



Université de Perpignan

ATTESTATION DE REUSSITE AU DIPLOME

Le Vice-Président Recherche atteste que

le Diplôme national de docteur BIOLOGIE
a été décerné à

Monsieur DAVID ROQUIS
né le 24 octobre 1982 à CERET (066)

au titre de l'année universitaire 2015/2016 avec la mention Très Honorable

Titre des travaux : Contribution de l'épigénétique dans les Dauernmodifikations et l'évolution adaptative chez le parasite humain Schistosoma mansoni et le corail tropical Pocillopora damicornis

Date de soutenance : 8 décembre 2015

Etablissement soutenance : UNIVERSITE DE PERPIGNAN

Jury : M. GUILLAUME MITTA, Président du jury, PROFESSEUR DES UNIVERSITES
UNIVERSITE DE PERPIGNAN
M. ETIENNE DANCHIN, Rapporteur du jury, DIRECTEUR DE RECHERCHE
UNIVERSITE TOULOUSE 3 PAUL SABATIER
M. RAYMOND PIERCE, Rapporteur du jury, DIRECTEUR DE RECHERCHE CNRS
Mme MARIE MIROUZE, Membre du jury, CHARGE DE RECHERCHE
M. OLIVIER REY, Membre du jury, CHARGE DE RECHERCHE
Mme CELINE COSSEAU, Co-Directeur, MAITRE DE CONFERENCE
UNIVERSITE DE PERPIGNAN
M. CHRISTOPH GRUNAU, Directeur de thèse, PROFESSEUR DES UNIVERSITES
UNIVERSITE DE PERPIGNAN

Ecole doctorale : Energie et Environnement

Fait à Perpignan, le 10 décembre 2015



Xavier PY

N° étudiant : 9011079

Avis important: Il ne peut être délivré qu'un seul exemplaire de cette attestation. Aucun duplicata ne sera fourni.

IX.6. ANNEXE 6 : LISTE DES CORRECTIONS DU MANUSCRIT DE THÈSE

Cette annexe dresse une liste des modifications apportées sur cette version post-soutenance du manuscrit (finalisée le 25 janvier 2016) par rapport à la version pré-soutenance (finalisée le 18 octobre 2015).

Modifications du texte

- **Remerciements** : ajout de quelques personnes malencontreusement oubliées ainsi que d'une section de remerciements post-soutenance.
- **I.1.3.7. Émergence de la génétique** (page 16-17) : ajout de la contribution de Wilhelm Johannsen.
- **I.1.3.9. Progrès moléculaires et géocentrisme** (page 18) : ajout de la définition du gène égoïste.
- **I.1.3.10 Importance des interactions** (pages 18-20) : approfondissement sur les travaux de William Donald Hamilton et ajout des contributions de George Price.
- **I.2.2.2. Autres modifications chimiques des nucléotides** (page 36) : cette section s'intitulait précédemment « autres nucléotides modifiés ».
- **I.2.2.5. Localisation nucléaire des chromosomes** (page 43-44) : approfondissement et ajout de références plus récentes.
- **I.3.1.1. Cycle de vie** (page 47) : ajout de détails sur le cycle de vie de *S. mansoni*.
- **I.3.5.2. Résistance induite à un schistosomicide** (page 57) : ajouts d'informations sur deux autres types d'induction de résistance à l'hycanthone.
- **I.4.1.1. Caractéristiques biologiques et cycle de vie** (pages 58-60) : ajouts de précisions sur les caractéristiques morphologiques des colonies de *P. damicornis*.
- **I.4.1.2. Distribution géographique** (page 60) : distribution géographique de *P. damicornis* mieux détaillée.
- **Chapitre VI : Discussion** (pages 203-218) : modifications, pour les trois types d'épimutations (cibées, aléatoires et associées au génotype), de leur nombre, de leur pourcentage d'héritabilité et de leurs caractéristiques générales pour être en accord avec la version de l'article « Frequency and mitotic heritability of epimutations in *Schistosoma mansoni* » acceptée par le journal **Molecular Ecology** le 24 janvier 2016.
- **VI.4. Rôle évolutif de la variabilité épigénétique** (pages 213-218) : modification du texte pour accompagner les nouvelles **Figures 31 et 32**.
- **Chapitre VIII : Références bibliographiques** (pages 223-243) : ajouts de douze nouvelles références reliées aux précédents ajouts et corrections.
- Multiples coquilles, fautes et détails mineurs corrigés tout au long du manuscrit.

Modifications des figures

- **Figure 2 et 9** (page 9 et page 29) : modification des polices pour améliorer la lisibilité.
- **Figure 3** (page 11) : refonte de la **Figure 3A et 3B**, ajout de la **Figure 3C**.
- **Figure 5** (page 23) : refonte de la **Figure 5A** pour rajouter *Daphnia longicephala* et ajout de la **5B** présentant les formes intermédiaires de *Daphnia cucullata*.
- **Figure 8** (page 28) : figure de meilleure résolution et redressée.
- **Figure 17** (page 44) : ajout de la **Figure 17B**.
- **Figure 27** (page 66) : correction d'une coquille.
- **Figure 31** (page 216) : ajout de cette nouvelle figure pour proposer un modèle adaptatif épigénétique à dans changements de l'environnement modérés.
- **Figure 32** (page 217) : ajout de cette nouvelle figure pour proposer un modèle adaptatif épigénétique à dans changements de l'environnement forts.

Documents supplémentaires

- **Chapitre II : Caractérisation de la structure chromatinienne de *S. mansoni*** (pages 95-97) : ajout de la *Supplementary Figure 1* et des *Supplementary Tables 1 et 2* de l'article de **PLoS Neglected Tropical Diseases**.
- **Chapitre III : Résistance à un schistosomicide et impact sur l'épigénome de *S. mansoni*** (pages 117-119) : ajout des *Supplementary Tables 1 et 2* et de la *Supplementary Figure 1* de l'article de **Frontiers in Genetics**.
- **Chapitre IV : Effet de la souche d'hôte intermédiaire sur l'épigénome de *S. mansoni*** (pages 125-170) : ajout de la version corrigée de l'article acceptée par **Molecular Ecology** (en date du 24 janvier 2016). Ajout des *Supplementary Figures 1 à 6* et du contenu des *Supplementary Files 4, 5, 6 & 8*.
- **Annexe 4** (pages 275-288) : ajout de cette annexe incluant les diapositives de la soutenance du 8 décembre 2015.
- **Annexe 5** (pages 289-304) : ajout de cette annexe incluant les différents documents administratifs reliés à la thèse.
- **Annexe 6** (pages 305-306) : ajout de cette annexe listant les modifications du manuscrit entre les versions pré- et post-soutenance.

Autre

- **Chapitre VIII : Références Bibliographiques** (pages 223-243) : refonte pour améliorer la lisibilité.

« Toute chose a une fin, sauf la banane, qui en a deux. »

Proverbe africain



Résumé

L'origine de la variabilité phénotypique est un sujet très débattu depuis les théories de Lamarck et Darwin. Dans la vision contemporaine de l'évolution adaptative, il est communément admis que la seule source héritable de variabilité phénotypique soit d'origine génétique. Le phénotype est alors le produit du génotype sous l'influence de l'environnement. La mutation aléatoire des séquences d'ADN permet de générer de nouveaux variants phénotypiques qui sont alors soumis à la sélection naturelle. Traditionnellement, il est considéré que les caractères acquis par un individu durant sa vie, en réponse à l'environnement, ne sont pas héréditaires et ne jouent aucun rôle évolutif. Pourtant, il y a presque un siècle, un biologiste allemand du nom de Victor Jollos a mis en évidence que certains phénotypes peuvent être induits par des conditions environnementales particulières et persister durant quelques générations en l'absence du stimulus initial avant de disparaître progressivement. Il nomma ce phénomène *Dauermodifikationen*, littéralement « modifications de longue durée ». Ses conclusions allaient à contre-courant des conceptions évolutives de son temps, et ont été considérées comme des artefacts expérimentaux. Toutefois, nous sommes maintenant conscient qu'outre le code génétique, il existe également un autre mécanisme permettant une réponse héréditaire et pourtant flexible en réponse aux fluctuations environnementales : le code épigénétique. Au cours de cette thèse, j'ai essayé de mieux caractériser le rôle des mécanismes épigénétiques, plus précisément ceux impliqués dans la structure chromatinienne, chez deux organismes présentant des *Dauermodifikationen* : le corail tropical *Pocillopora damicornis* et le parasite humain *Schistosoma mansoni*. Les deux objectifs principaux de cette étude sont de déterminer (I) de quelle manière l'environnement influence la structure chromatinienne (ciblée ou aléatoire) et (II) dans quelle mesure ces changements sont-ils héréditaires (mitotiquement ou méiotiquement).

Nos résultats ont permis de mieux caractériser les épigénomes des deux organismes étudiés. Nous avons décrit la structure chromatinienne de *S. mansoni* au travers de la distribution de six modifications d'histones, sur deux stades développementaux. Par ailleurs, nous avons montré chez *S. mansoni* trois types de changements de la structure chromatinienne : (I) ciblés en réponse à l'environnement, (II) associés au génotype et (III) aléatoires. Seuls les types II et III sont héréditaires d'un stade développemental du parasite à un autre. Nos travaux sur *P. damicornis* ont permis de remarquer une structure chromatinienne inhabituelle et d'offrir une première description d'un méthylome de corail.

Mots-clefs : *Dauermodifikationen*, épigénétique, modification d'histones, méthylation de l'ADN, *Schistosoma mansoni*, *Pocillopora damicornis*, évolution adaptative, hérédité non génétique.

Abstract

The origin of phenotypic variability has been much debated since the establishment of Lamarck's and Darwin's theories of evolution. It is commonly accepted in the contemporary vision of adaptive evolution that the only source of heritable phenotypic variability is genetic. Here, phenotypes are the product of the genotypes under the influence of the environment. Random DNA mutations generate novel phenotypes, which are then subjected to natural selection. Traditionally, it is considered that acquired characters are not heritable and have no impact on evolution. Yet almost a century ago, a German biologist named Victor Jollos revealed that some phenotypes could be produced in particular environmental conditions and could persist for a few generations in the absence of the original stimulus, before disappearing gradually. He named this phenomenon *Dauermodifikationen*, literally "long term changes". His conclusions were going against evolutionary conceptions of his time, and were considered experimental artefacts. However, we are now aware that, in addition to the genetic code, there is also another heritable, and yet flexible, mechanism responding to environmental fluctuations: the epigenetic code. In this thesis, I attempted to characterize the role of epigenetic mechanisms, and more specifically modifications of the chromatin structure, in two organisms with *Dauermodifikationen*: the tropical coral *Pocillopora damicornis* and the human parasite *Schistosoma mansoni*. The two main objectives of this study were (I) to determine how the environment influences the chromatin structure (in a targeted or random fashion) and (II) to what extent these changes are heritable (through mitosis or meiosis).

My results provide a better knowledge of the epigenome of the two organisms we studied. We have described the chromatin structure of *S. mansoni* through the distribution of six histone modifications, in two developmental stages. Furthermore, we have shown three types of changes in chromatin structure of *S. mansoni*: (I) targeted in response to environmental changes, (II) genotype associated, and (III) random. Only types II and III are inherited to the next developmental stages of the parasite. Our work on *P. damicornis* delivers evidence for an unusual chromatin structure in this organism and to provide the first description of a coral methylome.

Keywords: *Dauermodifikationen*, epigenetics, histone modifications, DNA methylation, *Schistosoma mansoni*, *Pocillopora damicornis*, adaptive evolution, non-genetic inheritance.